PB82-109604

Performance of Activated Sludge Processes Reliability, Stability and Variability

California Univ., Davis

Prepared for

Municipal Environmental Research Lab.
Cincinnati, OH

Sep 81

U.S. Department of Commerce
National Technical Information Service
NTIS

PB82-109604

EPA-600/2-81-227
September 1981

PERFORMANCE OF ACTIVATED SLUDGE PROCESSES
Reliability, Stability and Variability

by

Salar Niku
Edward D. Schroeder
George Tchobanoglous
Francisco J. Samaniego

Civil Engineering Department
University of California
Davis, California 95616

EPA Grant No. R805097-01

Project Officer
Jon Bender
Wastewater Research Division
Municipal Environmental Research Laboratory
Cincinnati, Ohio 45268

MUNICIPAL ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/2-81- 227	2. ORD Report	3. RECIPIENT'S ACCESSION NO. PB82 109604
4. TITLE AND SUBTITLE Performance of Activated Sludge Processes: Reliability, Stability, Variability		5. REPORT DATE September 1981 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Salar Niku, Edward D. Schroeder, George Tchobanoglous, and Francisco J. Samaniego		8. PERFORMING ORGANIZATION REPORT NO. Not Applicable
9. PERFORMING ORGANIZATION NAME AND ADDRESS Civil Engineering Department University of California Davis, California 95616		10. PROGRAM ELEMENT NO. P.E. 36B1C, Task C1 11. CONTRACT/GRANT NO. Grant No. R805097-01
12. SPONSORING AGENCY NAME AND ADDRESS Municipal Environmental Research Laboratory - Cinti,OH Office of Research and Development U.S. Environmental Protection Agency Cincinnati, Ohio 45268		13. TYPE OF REPORT AND PERIOD COVERED Final, 1979 14. SPONSORING AGENCY CODE EPA/600/14

15. SUPPLEMENTARY NOTES

Project Officer: Jon H. Bender (513) 684-7607

16. ABSTRACT

The objective of this research study was to statistically analyze the performance of a large number of activated sludge processes and to develop methods and procedures for introducing reliability and stability concepts into design and operation of treatment plants. The lognormal distribution was found to fit the observed effluent BOD and SS data more consistently. Using the lognormal distribution and a coefficient of reliability (COR), a probabilistic model and design tables and graphs have been developed for predicting achievable effluent BOD and SS concentrations. The proposed model can be used in design of a treatment process expected to perform at a certain reliability and/or to estimate the reliability of an operating treatment plant. The stability of various activated sludge processes was examined using several statistical measures. The standard deviation was found to be the most appropriate indicator of stability. Plants with effluent values below 10 g/m^3 are considered stable; plants with values greater than 10 g/m^3 are considered unstable. Correlations and regression analyses have been used to investigate the causes of effluent quality variability and to identify the extent of the factors contributing to this variability. No single or group of variables was found that could be used to characterize the variability of effluent quality for all plants in general. The use of geometric mean, as a measure of central tendency of daily effluent quality data, was recommended for setting discharge standards. An approach is presented that can be used to design a process stochastically when the effluent standards are deterministic in nature.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Sewage Treatment Statistical Analyses Activated Sludge Process Effluents Reliability Stability Variability		13B

18. DISTRIBUTION STATEMENT Release to the Public	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 165
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

DISCLAIMER

This report has been reviewed by the Municipal Environmental Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWARD

The U.S. Environmental Protection Agency was created because of increasing public and government concern about the dangers of pollution to the health and welfare of the American people. Noxious air, foul water, and spoiled land are tragic testimonies to the deterioration of our natural environment. The complexity of that environment and the interplay of its components require a concentrated and integrated attack on the problem.

Research and development is that necessary first step in problem solution; it involves defining the problem, measuring its impact, and searching for solutions. The Municipal Environmental Research Laboratory develops new and improved technology and systems to prevent, treat, and manage wastewater and solid and hazardous waste pollutant discharges from municipal and community sources, to preserve and treat public drinking water supplies, and to minimize the adverse economic, social, health, and aesthetic effects of pollution. This publication is one of the products of that research and provides a most vital communications link between the researcher and the user community.

Results of the statistical analysis of the variability in effluent quality of 43 activated sludge processes located throughout the United States are presented in this report. Daily composite data for one years operation was used in each case, and relationships between perforance characteristics and process loading and operation were developed.

Francis T. Mayo, Director
Municipal Environmental Research Laboratory

ABSTRACT

The objective of this research study was to analyze statistically the performance of a large number of activated sludge processes and to develop methods and procedures for introducing reliability and stability concepts into design and operation of treatment plants. First, variations in the daily effluent quality of 43 activated sludge processes were examined. From an analysis of the data, it was found that the effluent variables are not distributed symmetrically, and generally their distributions are skewed further to the right than to the left of the most frequent value. The lognormal distribution was found to fit to observed effluent BOD and SS data most consistently.

Using the lognormal distribution and a coefficient of reliability (COR), a probabilistic model and design tables and graphs have been developed for predicting achievable effluent BOD and SS concentrations. The COR is used to correlate the mean constituent value to a selected standard. The proposed model can be used in design of a treatment process expected to perform at a certain reliability and/or to estimate the reliability of a treatment plant under operation. The validity of the presented reliability model in prediction of effluent quality has been verified.

The stability of various activated sludge processes was examined using several statistical measures. Stability is defined as a measure of the daily variations from the annual mean. The standard deviation was found to be the most appropriate indicator of stability. A standard deviation of 10 g/m^3 was selected as stability cut-off point. Plants with effluent standard deviation values below 10 g/m^3 are considered stable; plants with values greater than 10 g/m^3 are considered unstable. The effect of plant size and process type on stability was also examined.

Correlation and regression analyses have been used to investigate the causes of effluent quality variability and to identify the extent of the factors contributing to this variability. The contribution of several variables to the observed effluent BOD and SS variations were examined. No single or group of variables was found that could be used to characterize the variability of effluent quality for all plants. However, the method may be applied to data from each individual plant to determine the principal factors responsible for effluent quality variations.

The appropriateness of effluent standards as they are currently stated was discussed, and the use of the geometric mean as a measure of central tendency of daily effluent quality data was recommended for setting discharge standards. This recommendation is based on the fact that effluent concentration values are distributed lognormally and the geometric mean of lognormal data has known probabilistic behavior and a meaningful interpretation.

Finally, application of study results to the design and the setting of discharge requirements was discussed. Design procedures and/or discharge requirements should be based on a recognition of the variability of process performance, that is, they should be stochastic in character. An approach was presented that can be used to design a process stochastically when the effluent standards are deterministic in nature.

This report was submitted in fulfillment of EPA-Grant No. R 805097-01 by the University of California under the sponsorship of the U.S. Environmental Protection Agency. The report covers the period April 1977 to September 1979, and work was finished as of September 1979.

CONTENTS

FOREWORD iii
ABSTRACT iv
CONTENTS vii
FIGURES x
TABLES xi
ABBREVIATIONS xiv
ACKNOWLEDGMENTS xvi

1. INTRODUCTION 1
    Objectives 2
    General Approach 2
    Data 3
    Note to the Reader 3
2. CONCLUSIONS AND RECOMMENDATIONS 6
    Conclusions 6
        Performance 6
        Reliability 6
        Stability 7
        Variability 7
    Recommendations 8
        Design 8
        Standards 9
3. STATISTICAL PARAMETERS MEASURING EFFLUENT QUALITY DATA . 10
    Objective 10
    Statistical Characterization of Effluent Quality 10
        Measure of Central Value 13
        Measure of Dispersion 13
        Measure of Asymmetry 14
        Discussion 14
    Effluent Concentration Distribution 14
4. RELIABILITY OF ACTIVATED SLUDGE PROCESSES 25
    Objective 25
    Reliability Concept 25
    Effluent Quality Distribution Characteristics 27
    Process Reliability 28
    Reliability Based Design of Activated Sludge Processes . . . 31
        Estimation of Reliability of Activated Sludge Processes . . . 31
        Graphical Form of the Reliability Model 44
    Verification of the Results 44

5. STABILITY OF ACTIVATED SLUDGE PROCESSES

Process Stability 45
Stability Versus Reliability 45
Background 45
Objective 46
Stability Measure 48
Selection of the Best Indicator of Stability 48
  Comment 50
  The Effect of Outliers on Stability Measures 50
  Time Series Analysis and Stability 50
  Ranking Treatment Plants with Respect to 54
    Their Stability
Stability Cut-Off Point 54
The Effect of Plant Size on Stability 60
The Effect of Process Type on Stability 63
  Statistical Examination of the Effect of Process . . . 63
    Type on Stability
  Discussion 64
The Effect of Process Type on Reliability 66
Stability Based Design of Activated Sludge Processes . . 66
6. FACTORS AFFECTING EFFLUENT VARIABILITY FROM 66
  ACTIVATED SLUDGE PROCESSES
Objective 69
Nature and Sources of Effluent Variability 69
  Influent Variables 69
  Environmental Conditions 70
  Biological and Operational Parameters 70
  Settling Characteristics and Clarifier Design 71
  Human Factors and Variability 72
  Inherent Variability 72
Statistical Analysis of Effluent Variations 73
  Descriptive Statistics 73
  Discussion 74
Correlation Analysis 83
  Linear Correlation 83
  Non-Linear Correlation 84
  The Effect of Lag Periods 87
  The Effect of Outliers 87
Regression Analysis; Examination of the Effect of 90
  Selected Variables on Effluent Quality Variations
  Discussion 90
  Process Control 100
7. ON SETTING STANDARDS FOR WATER QUALITY MANAGEMENT . . . 101
Objective 103
Proposal of Geometric Mean as a Measure of Central . . . 103
  Tendency of Effluent Concentration
  Measuring Central Tendency 103
  Discussion 104
Water Quality Standards Based on Geometric Mean 104
8. DISCUSSION 106
Facts 108
Stochastic Based Design and Standards 108
. . . . 109

Design Procedure 110
Operational Application 111
Standards 111
Reliability Estimation and the Use of the Geometric Mean 112
Process Control 114

REFERENCES 116

APPENDICES

A. Scattergrams of Daily Effluent BOD and SS Data 120
B. Examination of the Distribution of 7-Day and 30-Day Averages 143
C. Lognormal Probability Distribution. 146
D. Kolmogorov-Smirnov goodness-of-fit test. 148

FIGURES

Number **Page**

1 Histogram and Probability Density Function of Observed BOD Concentration 15

2 Histogram and Probability Density Function of Observed SS Concentration 16

3 Probability Plot for Effluent BOD Concentration 17

4 Probability Plot for Effluent SS Concentration 18

5 Probability Plot for Logarithm of Effluent BOD Concentration 20

6 Probability Plot for Logarithm of Effluent SS Concentration 21

7 Reliability Versus Normalized Mean (COR) for Different Coefficient of Variations 34

8 Percentile Values Versus Normalized Variates for Effluent BOD and SS Concentration 43

9 Reliability Versus Stability 47

10 Changes in Probability Density Function (PDF) with Changes in Means and Standard Deviations 49

11 Mean, Standard Deviation, and Range of Effluent BOD Concentration in Different Treatment Plants. 61

12 Mean, Standard Deviation, and Range of Effluent SS Concentration in Different Treatment Plants. 62

13 Effluent BOD as a Function of Mean Cell Residence Time 82

TABLES

Number **Page**

1 Activated Sludge Treatment Plants Studied 4

2 Statistics of Secondary Effluent BOD and SS Concentration 11

3 Effluent BOD and SS Distributions: Kolmogorov-Smirnov Goodness-Of-Fit Test 23

4 Cumulative Probabilities of the Standard Normal Distribution 30

5 Coefficient of Reliability (COR) as a Function of V_X and Reliability Based on Lognormal Distribution 32

6 Reliability as a Function of Normalized Mean and V_X of Effluent Concentration 33

7 Percentile Values of Effluent BOD Concentration 35

8 Percentile Values of Effluent SS Concentration 37

9 Coefficient of Reliability as a Function of V_X and Reliability from Pooled Plant Data 40

10 Comparison of Predicted and Measured Performance of Activated Sludge Plants Across the United States 41

11 Pearson Correlation Coefficients, r, of Annual Data for 43 Activated Sludge Treatment Plants 51

12 Statistics of 99%ile Data of Effluent BOD and SS Concentration 53

13 Estimatation of Standard Deviation by Mean Square Error, ($S_x = \sqrt{MSE}$), of Best Fitted Line Through Effluent BOD and SS Concentrations 55

14 Ranking of Wastewater Treatment Plants with Respect to Their Stability, Based on the Standard Deviation of 100%ile and 99%ile, Mean Square Error, and Coefficient of Stability of Effluent BOD Concentration 56

15 Ranking of Wastewater Treatment Plants with Respect to Their Stability, Based on the Standard Deviation of 100%ile and 99%ile, Mean Square Error, and Coefficient of Stability of Effluent SS Concentration 58

16 Statistics of the Annual Effluent BOD and SS Concentration Data for Different Process Types 65

17 Estimated Reliability of Various Process Types, Assuming Standard Level of $X_s = 30\ g/m^3$ 67

18 Recommended Design Values for Effluent BOD and SS Concentrations in Order to Produce Stable Process 68

19 Statistics of Some Selected Variables for Different Treatment Plants 75

20 Correlation Analysis; Pearson Correlation Coefficient, r, Between Effluent BOD and Selected Variables 85

21 Correlation Analysis; Pearson Correlation Coefficient, r, Between Effluent SS and Selected Variables 86

22 Non-Linear Correlation Analysis; Pearson Correlation Coefficient, r, Between Effluent BOD and Selected Variables 88

23 Non-Linear Correlation Analysis; Pearson Correlation Coefficient, r, Between Effluent SS and Selected Variables 89

24 Correlation Analysis with Lags; Pearson Correlation Coefficient, r, Between Effluent BOD and Selected Variables 91

25 Correlation Analysis with Lags; Pearson Correlation Coefficient, r, Between Effluent SS and Selected Variables 93

26 Correlation Analysis of 99%ile of Data; Pearson Correlation Coefficent, r, Between Effluent BOD and 99%ile Data of Selected Variables 95

27 Correlation Analysis of 99%ile of Data; Pearson Correlation Coefficient, r, Between Effluent SS and 99%ile Data of Selected Variables 96

28 Regression Analysis; Coefficient of Determination, R , for Effluent BOD and Selected Groups of Variables. 98

29 Regression Analysis; Coefficient of Determination, R , for Effluent SS and Selected Groups of Variables. 99

30 Number of Times That Each Variable was Contributed Significantly (Greater than 5 Percent) To Effluent Quality Variations of 43 Plants 102

31 The Influence of Outliers on Three Measures of Central Tendency 105

ABBREVIATIONS

BOD	5-day biochemical oxygen demand
BOD_i	influent BOD
COR	coefficient of reliability
COS	coefficient of stability
EPA	U.S. Environmental Protection Agency
exp	exponential
F:M	food to microorganism ratio
f_X	probability density function of X
GM	geometric mean
g/m^3	gram per cubic meter
K_T and K_{20}	removal rate coefficients at T and 20°C, respectively
Ln	natural logarithm
MCRT	mean cell residence time
mgd	million gallons per day
MLSS	mixed liquor suspended solids
MSE	mean square error
m_X	mean of X
m_X	median of X
m_{LnX}	mean of the natural logarithm of X
P	probability
Q_R	return sludge flow
Q_W	wastage sludge flow rate
r	bivariate correlation coefficient
r^2	bivariate coefficient of determination
R^2	multiple coefficient of determination
SETSOL	settleable solids
SS	suspended solids
SS_i	influent SS
S_X	sample standard deviation
SVI	sludge volume index
X_i	observed variable concentration
X	sample arithmetic mean
X_R	return sludge concentration
X_s	a fixed standard (e.g., 30 g/m^3)
V_X	sample coefficient of variation of X
WTEMP	wastewater temperature
Z	standarized normal variable
$Z_{1-\alpha}$	standardized normal variate at 1-α degree of probability
α	probability of failure
σ_X	standard deviation of X

σ_X^2	variance of X
σ_{LnX}	standard deviation of natural logarithm of X

ACKNOWLEDGMENTS

The cooperation of the U.S. Environmental Protection Agency, various water pollution control agencies, and the staffs of the treatment plants who generously provided us with the treatment plant data is appreciated; without their cooperation this work could not have been completed. Support by the project officer, Mr. Jon Bender, was particularly appreciated. The assistance provided by Mr. Richard S. Haugh, Department of Civil Engineering, Mr. Harold K. Grade, Programming Consultant, is acknowledged gratefully. We also with to thank Ms. Marlene Pendergrass and Ms. Carol Hoffman for their careful typing.

This report was originally submitted by Dr. Salar Niku as a doctoral dissertation in the Department of Civil Engineering at the University of California, Davis.

SECTION 1

INTRODUCTION

In recent years regulatory agencies have become more aware of the need to examine continually the performance characteristics of wastewater treatment plants to respond to the growing interest of the Nation on water quality and pollution control. Since 1970, explicit performance standards have been placed on most biological processes in the United States. The Federal Water Pollution Control Act Amendments of 1972 (P.L. 92-500) were written with the intent of restoring and maintaining the chemical, physical, and biological integrity of the Nation's water (17). To achieve these objectives, the U.S. Environmental Protection Agency (EPA) has set a minimum standard of effluent quality as that attainable by the application of biological treatment processes. (12) The standards for secondary treatment were to be met by 1977, and Best Practicable Waste Treatment Technology is to be achieved by 1983. Comprehensive standards have not been established because of the lack of information on the performance capabilities of operating plants.

As a result of the passage of the 1972 Water Pollution Control Act, the design and construction of wastewater treatment plants has flourished throughout the United States. Currently, activated sludge processes are the most commonly used method for treating municipal wastewaters. Although this process has been used for more than 60 years, little is known about its performance characteristics. Until the late 1950's, process design involved the application of a set of empirical guidelines. These guidelines were indirectly related to process performance, but they did not provide for prediction of performance. The introduction of stoichiometric and kinetic concepts of process design has resulted in some ability to predict soluble effluent BOD values. Estimates of effluent suspended solids are still based on experience. In addition, in the application of the kinetic and stoichiometric procedures used in the design of municipal and industrial wastewater processes, steady-state conditions are assumed. These processes are rarely operated under steady-state conditions, however. Fluctuations in input loads (flow, BOD and SS concentration), environmental conditions (temperature), and in-plant biological and operational variations (recycle rate, wastage sludge rate), are given little consideration in the system design. Unit processes and operations are sized on the basis of average or peak flows, but in each case steady-state conditions are assumed.

Presently, very little is known about the actual on line performance of wastewater treatment processes. Process models currently available (25,28,34) are based on mean values and cannot be used to predict variations in performance. Thus both the designer choosing processes and the regulatory agency staff member evaluating project proposals and setting discharge requirements have little information to use in their respective decision making functions. The lack of precise design methods, uncertainties, and the dynamic nature of biological waste treatment processes have also led designers to introduce oversized units and to overdesign processes. Paradoxically, economic pressure and/or lack of understanding of the variables that affect effluent quality has also resulted in inadequate or undersized processes in many cases.

OBJECTIVES

The basic purpose of this study is to develop information that can (1) be used by the designer and operator of a wastewater treatment plant in evaluating and predicting the performance of a biological treatment process with respect to effluent quality and (2) by the regulatory agencies in preparing reasonable, effective, and technically sound discharge requirements.

No wide ranging analyses have been conducted of process data or studies that specifically address questions such as: 1) How reliable are the present secondary treatment processes and what performance level can they be expected to achieve? 2) Are secondary treatment process stable? 3) Is there any relationship between the type of process and stability? 4) What are the nature and extent of effluent variations? and 5) Are the present standards structured in the most appropriate manner? To answer such questions, three subjects will be studied: Reliability, stability, and variability of secondary treatment process effluent quality.

Study of the reliability of wastewater treatment processes will require developing a reliability model that can be used to predict the process performance of plants under design or currently under operation. Stability of wastewater treatment plants will be examined quantitatively with the objective of identifying those processes that are unstable and to identify some possible reasons for the instabilities. Certainly a more stable process requires less operative attention and is more desirable. The fourth question will be addressed by examining the variability of effluent quality and by evaluating the sensitivity of effluent quality to selected process variables and parameters. Improved understanding of process performance variability will be of use in designing and operating of plants with decreased effluent quality variability. Decreased variability will allow process designers to improve system performance and should result in significant decreases in the pollution load on receiving waters and in the reduced operational costs. Finally, the fifth question will be addressed, and the appropriateness of current effluent standards will be examined. An alternative manner of stating effluent standards is presented.

GENERAL APPROACH

Establishing a quantitative approach for the examination of the performance of the treatment processes required statistical analysis of data obtained from plant operation. The scope of this project included collection and analysis of performance and operational data from a large number of activated sludge wastewater treatment plants located throughout the United States. Evaluation of the performance of these processes in terms of reliability stability, and variability was based on the analysis of these data.

Analysis was carried out at two levels. In the first level, the plant was considered as a "black box" with time-varying outputs. Input variables were considered to be uncontrolled and unknown. Effort was made to develop reliability and stability models from such data that can be used to evaluate and predict process performance and that can be useful to regulatory agencies and to designers. The "black box" level was utilized in Sections 3, 4, and 5.

In the second level, the effects of the input loadings, and environmental and operating parameters were considered. This level was developed in Section 6. The nature, causes, and the extent of effluent variability was discussed, and effort was made to identify factors contributing to this variability. The knowledge of variations gained in this study will aid in discovering ways to improve the stability and reliability of biological wastewater treatment processes.

This study was generally based on statistical and probability techniques. The reader is referred to explanations of the principles of statistical procedures in the references at the end of the paper. (4,8,11,27,29,30)

DATA

Requests were made for the daily operating data from more than 600 treatment plants across the United States. Approximately 100 plants responded to the request; daily data were provided for about 50 plants. Forty-three plants having the most complete sets of data were used in this study. Location by state, process type, and size of each plant used are shown in Table 1. In one plant, five years of data were used. Each operating year was considered as a separate plant (plants number 24 to 28). Plants numbered 5, 6, 7; 14, 15; 20 to 23; 29, 30; and 42, 43 are actually "batteries" of treatment plants. Each battery was operated as a separate independent plant. Generally, each plant had one complete fiscal year of data, however, in some plants data were derived from two fiscal years. As long as one continuous year of data were available these data were used in the study.

The selected treatment plants represent a range of geographic locations and environmental conditions throughout the United States. Plants range in flow from less than one mgd to more than 200 mgd and include municipalities with no to a large percentage of industrial wastes. Activated sludge process modifications included in the study were conventional activated sludge, step-feed and step aeration, contact-stabilization, complete-mixed, kraus, and extended aeration.

Computer programs were developed to compile and analyze the data. Different available statistical packages were studied, and the Statistical Package for the Social Sciences (30) (SPSS) was chosen to use throughout the study primarily because of the package's potential for handling missing data and for making large files.

NOTE TO THE READER

The purpose of this study was not to evaluate the performance of individual activated sludge treatment plants, but to select a representative group of such treatment plants and use daily data from these plants to ascertain some general

TABLE 1. ACTIVATED SLUDGE TREATMENT PLANTS STUDIED

Plant number	Plant location	Year	Type of process	Average flow (mgd)
1	Michigan	76	Complete Mix	14.2
2	California	07-75/6-76	Conventional	100.0
3	California	76	Complete Mix/Contact Stab.	16.2
4	Wisconsin	76	Contact Stabilization	23.7
5	Illinois	76	Step Feed	69.7
6	Illinois	76	Conventional	66.2
7	Illinois	76	Conventional	69.1
8	Maryland	76	Step Feed	15.2
9	New York	77	Complete Mix	18.6
10	Alabama	06-76/5-77	Aerated Lagoon	23.6
11	Michigan	10-76/9-77	Complete Mix	0.74
12	Virginia	76	Coventional/Step Feed	15.8
13	Alabama	07-76/6-77	Conventional/Step Feed	10.5
14	Illinois	76	Conventional	2.2
15	Illinois	76	Step Feed	3.3
16	California	75	Step Aeration	6.3
17	California	76	Step Aeration	2.4
18	California	76	Step Aeration	29.0
19	California	76	Step Aeration	7.1
20	Illinois	76	Conventional	209.0
21	Illinois	76	Conventional	189.0
22	Illinois	76	Conventional	188.0
23	Illinois	76	Step Feed	185.0
24	Michigan	73	Conventional	1.9
25	Michigan	74	Conventional	2.1
26	Michigan	75	Conventional	2.1
27	Michigan	76	Conventional	2.3
28	Michigan	77	Conventional	2.1
29	Michigan	77	Conventional	9.5
30	Michigan	77	Conventional	9.5
31	Connecticut	76	Conventional	3.8
32	California	76	Plug Flow/Kraus	47.1
33	Ohio	76	Contact Stabilization	2.8
34	Nevada	75	Conventional/Tapered	17.0
35	Kansas	76	Contact Stabilization	10.4
36	Washington	76	Conventional/Step Feed	30.8
37	California	76	Complete Mix/Step Feed	13.9
38	Michigan	10-76/10-77	Step Feed	17.3
39	Michigan	77	Conventional/Contact Stab.	33.3
40	Michigan	12-76/11-77	Extended Aeration	0.56
41	Michigan	01-77/10-77	Conventional	26.9
42	Wisconsin	76	Conventional	81.6
43	Wisconsin	76	Conventional	54.0

stochastic rules and facts that are common to all the plants. Therefore, any data presented or any statement made in reference to a particular plant in this paper serves only as an example; a judgement concerning the operation or design of that plant is not intended.

Over 15,500 days of plant operational data were compiled and analyzed in this study. The extent of computer printouts is so great that it makes it impossible to put the results into a readable format. Therefore, the data and most of the detailed results are not presented here.

SECTION 2

CONCLUSIONS AND RECOMMENDATIONS

The daily performance records of 43 activated sludge wastewater treatment processes have been analyzed statistically to examine the natural variations in effluent quality and to develop reliability and stability concepts for the design and operation of treatment plants. The selected treatment plants represent different modifications of the activated sludge process, and range in flow from less than one mgd to more than 200 mgd.

CONCLUSIONS

Conclusions derived from this study are presented in four categories: performance, reliability, stability, and variability.

Performance

1. Effluent concentrations of BOD and SS varied significantly amongst the plants in this study. Mean effluent BOD values ranged from 1.9 to 85.6 g/m^3 with a group average of 15.8 g/m^3. Mean values for effluent SS ranged from 2.5 to 80.1 g/m^3 with a group average of 19.4 g/m^3.

2. On the average, the effluent SS concentration and its variation is greater than the corresponding effluent BOD values. Solids separation facilities are presently a limiting factor in the performance of activated sludge treatment systems.

3. Effluent concentration distributions are not symmetrical and generally are skewed further to the right than to the left of the most frequent value.

4. A lognormal distribution consistently was found to fit the observed effluent BOD and SS data best.

Reliability

5. A probabilistic approach to process design provides a theoretical basis for the analysis of process reliability.

6. The Coefficient of Reliability (COR), defined on a probability basis, can be used to relate mean constituent values to selected discharge standards.

7. A reliability model has been developed that can be used for both the estimation of the mean value in design of a process or for the prediction of effluent quality in a plant under operation.

Stability

8. "Standard deviation" is the most appropriate measure of stability.

9. Plants having a standard deviation of less than 10 g/m^3 for both effluent BOD and SS may be considered as "stable" statistically; plants with a standard deviation of greater than 10 g/m^3 may be considered "unstable."

10. The mean effluent BOD for stable plants is usually less than 17 g/m^3; the mean effluent SS values is less than 15 g/m^3. Effluent BOD and SS concentrations in stable plants rarely exceeds 60 g/m^3.

11. Unstable plants have annual mean effluent BOD and SS values greater than 16 g/m^3 and 10 g/m^3, respectively. The maximum daily BOD and SS concentrations of these plants are usually much greater than stable plants and exceeds 90 g/m^3 and 130 g/m^3, respectively.

12. Process upsets occur more often due to high values of SS as opposed to high values of BOD in the final effluent.

13. There is no relationship between the size of plant and stability.

14. Various activated sludge processes ranked in terms of their stability and reliability are as follows: 1) step-feed and step aeration, 2) conventional, 3) complete-mix, and 4) contact-stabilization.

Variability

15. Generally, no single variable can be used to characterize the variability of effluent BOD and SS in all plants. Coefficients of correlation between effluent concentration and other variables are insignificant and vary in both strength and sign.

16. No significant time related trends or cycles were found for activated sludge effluent data.

17. The effect of the lag period between influent and effluent variables is insignificant.

18. The combination of all available variables (input, environmental, and operational) can only be used to explain between 13 and 61 percent of the effluent BOD and between 5 and 52 percent of the effluent SS fluctuations (depending on

the plant). The remaining variations (which may categorized as being due to inherent or human factors) are unexplained.

19. Variables controlling plant performance vary from plant to plant. In most cases, one or two variables can be used to explain more than 90 percent of the total variability that was accountable.

20. Among all variables, wastewater temperature was the most significant in accounting for effluent variations. Other variables found to be significant (minimum significance of 5 percent) depending on the plant include time, flow, SVI, influent BOD, MCRT, MLSS, and influent SS.

21. In many plants, the MCRT was not properly used for plant operation. Only a few plants were operated in such a way that the MCRT was a useful operating parameter.

RECOMMENDATIONS

Recommendations based on the results of this study are divided into two categories dealing with 1) the design of treatment processes and 2) the setting of discharge requirements.

Design

1. The uncertainties associated with plant design and operation should be recognized in the design process. The best way to accomplish this is to incorporate stochastic concepts and procedures in the design process.

2. The mean value for the design of activated sludge processes may be estimated using the coefficient of reliability (COR) developed in this study.

3. The reliability model developed in this study may be used to estimate the reliability of a treatment plant (i.e., the fraction of time that the effluent concentration exceeds any standard value).

4. To produce a stable effluent quality (i.e., a process with a standard deviation of equal to or less than 10 g/m^3), the process should be designed for a mean BOD and SS value equal to or less than $m_X = S_X/V_X$, where S_X, the standard deviation, is equal to 10 g/m^3 and V_X, the coefficient of variation, must be estimated.

5. To achieve stability, the activated sludge process should be designed for an effluent BOD concentration equal to or less than 13 to 15 g/m^3 and an effluent SS concentration equal to or less than 10 to 12 g/m^3.

Standards

6. The geometric mean of the daily effluent quality data is a more appropriate measure of central tendency than the arithmetic mean and is recommended parameter to be used in setting standards in water quality management.

7. Current EPA effluent discharge requirements should be restructed on the basis of stochastic concepts to permit individual parameter values to exceed the designated standard value a given number of times in a specified time period.

SECTION 3

STATISTICAL ANALYSIS OF EFFLUENT QUALITY VARIATION

Effluent quality variations were analyzed statistically to develop preliminary information that could be used to define and study the reliability and stability of wastewater treatment plants. The "black box" approach was used in this portion of the study. In this level of analysis, input, operational, and environmental variables were assumed to be unknown. Extent and magnitude of variations in effluent quality were examined only by consideration of effluent performance data. Results of black box analysis do not provide any cause and effect relationships or implications. However, where consistent patterns of results or frequency distributions can be established, extremely important and useful information can be derived using the black box approach.

OBJECTIVE

It has been shown in many previous studies that the performance of activated sludge wastewater processes is highly variable. (1,3,15,18,21,33,35,44) In most cases, this variability has been examined qualitatively. The objective here was to examine the natural variations in effluent quality by analyzing statistically effluent quality data and to determine its distributional characteristics. Such an analysis has a direct bearing on water quality management since effluent quality variation results in variable loading on receiving waters and variability in downstream water quality. The results of this analysis may be used to determine the expected effluent quality of various activated sludge processes for design purposes and are also useful in the examination of stability and reliability of different plants.

STATISTICAL CHARACTERIZATION OF EFFLUENT QUALITY

Preliminary statistical analysis of the data was required to provide some initial insight or orientation into process performance. This included determination of statistical parameters such as the mean, standard deviation, skewness, maximum and minimum values for each data set. These parameters are summarized for the plants used in this study in Table 2, and explanation of the significance of each parameter follows.

TABLE 2. STATISTICS OF SECONDARY EFFLUENT BOD AND SS CONCENTRATION

Plant Number	BOD							SS						
	Mean, g/m^3	S_X, g/m^3	V_X	Skew	Max, g/m^3	Min, g/m^3	Valid Obs.	Mean, g/m^3	S_X, g/m^3	V_X	Skew	Max, g/m^3	Min, g/m^3	Valid Obs.
1	9.74	7.30	0.75	2.17	65.0	0.3	326	10.95	12.82	1.17	5.15	150.0	0.2	364
2	8.47	4.37	0.52	1.00	25.0	0	362	6.26	2.69	0.43	2.50	30.0	1.0	362
3	10.00	6.67	0.67	2.12	43.0	1.0	361	14.54	4.60	0.32	0.45	28.0	3.0	366
4	18.33	11.40	0.62	1.95	76.0	5.0	364	26.21	21.87	0.83	2.52	158.0	6.0	364
5	16.23	9.27	0.57	1.57	63.0	3.0	363	21.17	15.67	0.74	3.98	145.0	3.0	365
6	16.54	9.28	0.56	1.36	60.0	3.0	362	21.92	16.11	0.73	3.84	139.0	5.0	366
7	21.23	11.50	0.54	2.82	92.0	5.0	363	16.66	21.23	1.27	5.55	213.0	2.0	366
8	16.34	16.10	0.99	6.84	202.0	3.0	260	20.98	26.53	1.26	4.59	224.0	2.0	259
9	24.76	22.44	0.91	2.19	174.0	1.0	239	44.47	48.66	0.98	2.20	322.0	4.0	243
10	30.35	11.61	0.38	1.46	91.0	9.0	353	58.79	18.71	0.32	1.34	150.0	21.0	356
11	18.46	16.59	0.90	3.20	110.0	2.0	365	10.32	17.54	1.70	6.17	186.0	1.0	365
12	16.56	7.05	0.43	1.73	52.0	5.7	362	14.70	11.55	0.79	4.59	136.0	2.0	332
13	16.32	14.18	0.87	2.07	94.0	1.0	341	14.74	16.72	1.13	2.80	143.0	1.0	344
14	22.94	21.83	0.95	4.54	242.0	4.0	357	34.74	47.75	1.37	5.08	534.0	3.0	365
15	26.56	22.90	0.86	2.08	170.0	3.0	356	28.17	34.03	1.21	2.64	244.0	1.0	364
16	4.90	2.51	0.51	1.27	16.0	1.0	211	5.35	2.93	0.55	1.87	23.0	1.0	212
17	1.94	1.39	0.72	1.91	7.0	0.9	78	2.56	2.88	1.13	2.82	22.0	0.6	366
18	2.27	0.78	0.34	0.48	4.0	1.0	81	4.49	2.13	0.47	0.85	12.0	1.0	366
19	4.81	2.02	0.42	0.59	11.0	1.0	83	8.20	3.93	0.48	0.94	24.0	1.0	366
20	7.73	3.76	0.49	0.91	20.0	2.0	365	7.00	4.47	0.64	5.32	56.0	1.0	366
21	4.43	2.38	0.56	1.79	18.0	1.0	366	6.34	3.67	0.58	3.66	40.0	1.0	366
22	4.26	2.03	0.48	1.35	14.0	1.0	366	5.78	3.09	0.53	2.42	27.0	1.0	366
23	2.97	1.71	0.58	1.59	10.0	1.0	366	3.60	2.03	0.56	1.55	15.0	1.0	365
24	24.36	27.00	1.11	1.82	140.0	1.0	353	18.88	25.09	1.33	3.88	220.0	1.0	348
25	12.86	9.03	0.70	1.84	61.0	1.0	351	11.90	12.70	1.07	6.82	151.0	1.0	362
26	9.06	4.39	0.48	1.21	28.0	2.0	361	9.68	6.37	0.66	2.16	54.0	1.0	362

TABLE 2 (continued)

Plant Number	BOD							SS						
	Mean, g/m^3	S_X g/m^3	V_X	Skew	Max, g/m^3	Min, g/m^3	Valid Obs.	Mean, g/m^3	S_X g/m^3	V_X	Skew	Max, g/m^3	Min, g/m^3	Valid Obs.
27	12.57	6.47	0.51	0.85	38.0	1.0	362	8.89	5.51	0.62	1.17	30.0	1.0	365
28	13.77	6.58	0.48	0.87	48.0	2.0	362	8.98	7.20	0.80	2.62	62.0	1.0	362
29	4.28	3.80	0.89	1.64	24.0	1.0	365	9.97	5.80	0.58	1.14	37.0	1.0	365
30	5.99	5.68	0.95	1.63	26.0	1.0	365	10.06	7.09	0.70	1.86	55.0	1.0	365
31	10.83	8.48	0.78	2.77	50.0	1.0	149	10.90	4.79	0.44	2.33	48.0	3.0	363
32	24.02	9.85	0.41	0.35	50.0	4.3	360	24.12	9.26	0.38	1.17	62.0	10.0	365
33	18.01	16.19	0.90	1.99	98.0	1.0	152	21.76	16.38	0.75	2.29	114.0	1.0	157
34	8.78	5.24	0.60	1.50	40.0	2.0	323	14.78	7.69	0.52	1.85	73.0	2.0	364
35	31.64	27.14	0.86	3.47	248.0	2.0	366	35.47	37.92	1.07	2.23	264.0	1.0	348
36	6.72	4.68	0.70	1.61	20.7	0.6	247	7.28	3.87	0.53	1.86	27.0	1.0	347
37	21.01	12.77	0.61	2.14	120.0	5.0	345	19.11	14.61	0.76	0.83	61.0	1.0	337
38	16.50	19.84	1.20	3.63	164.0	3.0	363	65.03	88.86	1.37	3.89	780.0	5.0	364
39	85.55	57.96	0.68	1.42	345.0	10.0	365	80.08	74.46	0.93	2.41	492.0	6.0	365
40	14.41	5.31	0.37	1.17	32.0	5.0	251	8.82	5.28	0.60	1.78	34.0	2.0	251
41	9.03	3.84	0.43	0.92	23.0	2.0	304	9.81	6.79	0.69	1.75	39.0	2.0	304
42	20.69	14.07	0.68	2.91	110.0	5.8	360	28.41	37.06	1.30	4.17	311.0	2.0	356
43	21.40	27.18	1.27	5.00	260.0	4.2	366	42.30	70.87	1.68	4.16	605.0	2.0	366

[a] All BOD values are 5-day BOD.

Measure of Central Value

The arithmetic mean value, X, is used to describe the central tendency of the data and may be regarded as a "typical" value representing a set of data. If the sequence of observed values is denoted $X_1, X_2, \ldots, X_n$, the mean X is simply (8)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad (1)$$

Mean effluent BOD values for the plants included in this study ranged between 1.9 to 85.6 g/m^3 with a group average of 15.8 g/m^3. Similar mean values for effluent SS ranged between 2.5 to 80.1 g/m^3 with a group average of 19.4 g/m^3. Levels of effluent BOD and SS concentrations differ greatly between plants.

Measure of Dispersion

Common measures of data dispersion are the variance, standard deviation, and coefficient of variation.

Variance and Standard Deviation--Variance, S_X^2, is defined as the weighted average of the squared deviations from the mean: (8)

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad (2)$$

When the data are clustered closely to the mean value, the variance is small; as the data become more spread out, variance increases. In practice, the positive square root of the variance, called "standard deviation", S_X, is given more importance than the variance because this value has the same unit value as the variable. Standard deviation can be compared easily with the mean to gain some feeling for the degree of uncertainty associated with the variable.

For the 43 treatment plants studied, standard deviations of effluent BOD and SS varied between 0.8 to 58 g/m^3 and 2.1 to 89 g/m^3, respectively. From this wide range of standard deviations, it can be concluded that some plants are subject to more variation than others with respect to effluent quality variability.

Coefficient of Variation--The coefficient of variation, V_X, is a measure of relative dispersion and is useful in comparing various data sets. This coefficient is defined as the ratio of the standard deviation to the mean. (8)

$$V_X = \frac{S_X}{\bar{X}} \qquad (3)$$

As shown in Table 2, the coefficient of variation for BOD ranged between 0.34 to 1.27 with an average value of 0.68 and a standard deviation of 0.23. A value of V_X greater than one was computed for three plants, and 25 percent of the plants had values of V_X less than 0.50. The coefficient of variation for effluent SS ranged between 0.32 to 1.70 with an average value of 0.84 and a standard deviation of 0.37. A value of V_X greater than one was computed for 14 plants, and 16 percent of the plants had values of V_X less than 0.50.

Measure of Asymmetry

The coefficient of skewness is a measure of deviation from symmetry. This coefficient is defined by (8)

$$\text{skewness} = \frac{1}{n} \sum_{i=1}^{n} [(X_i - \bar{X})/S_X]^3 \qquad (4)$$

This coefficient is close to zero when the distribution is a symmetric curve. The coefficient tends to be positive for distributions skewed to the right (i.e., the data are clustered more to the left of the mean with most of the extreme values to the right) and negative for those skewed to the left. For the 43 treatment plants studied, skewness coefficient was always positive, indicating that generally the data were skewed to the right.

Discussion

Based on the results summarized in Table 2, it can be concluded that the performance of these plants on an annual basis is quite variable. On the average, effluent SS concentration content and variation is somewhat greater than that of effluent BOD. On the basis of this finding, it appears that the solids separation facilities are probably a factor limiting the efficiency of activated sludge treatment systems.

EFFLUENT CONCENTRATION DISTRIBUTION

Effluent variability can be illustrated with a histogram and analyzed by determining the probability density function of the data. Typical histograms of the effluent BOD and SS of the plants studied for one year of data are shown in Figures 1 and 2. As shown, the data are generally not symmetrical and are skewed to the right.

The histogram may be approximated by a superimposed continuous smooth curve, a scaled version of which is called the probability density function. If the cumulative distribution function of the probability law plots as a straight line on a normal probability paper, then the data distribution is said to be normal. (8) The arithmetic probability plots of BOD and SS concentration for the same plant are shown in Figures 3 and 4. The nonnormality of these data is obvious. Thus, methods used to describe or predict results from data of this type should not assume normality.

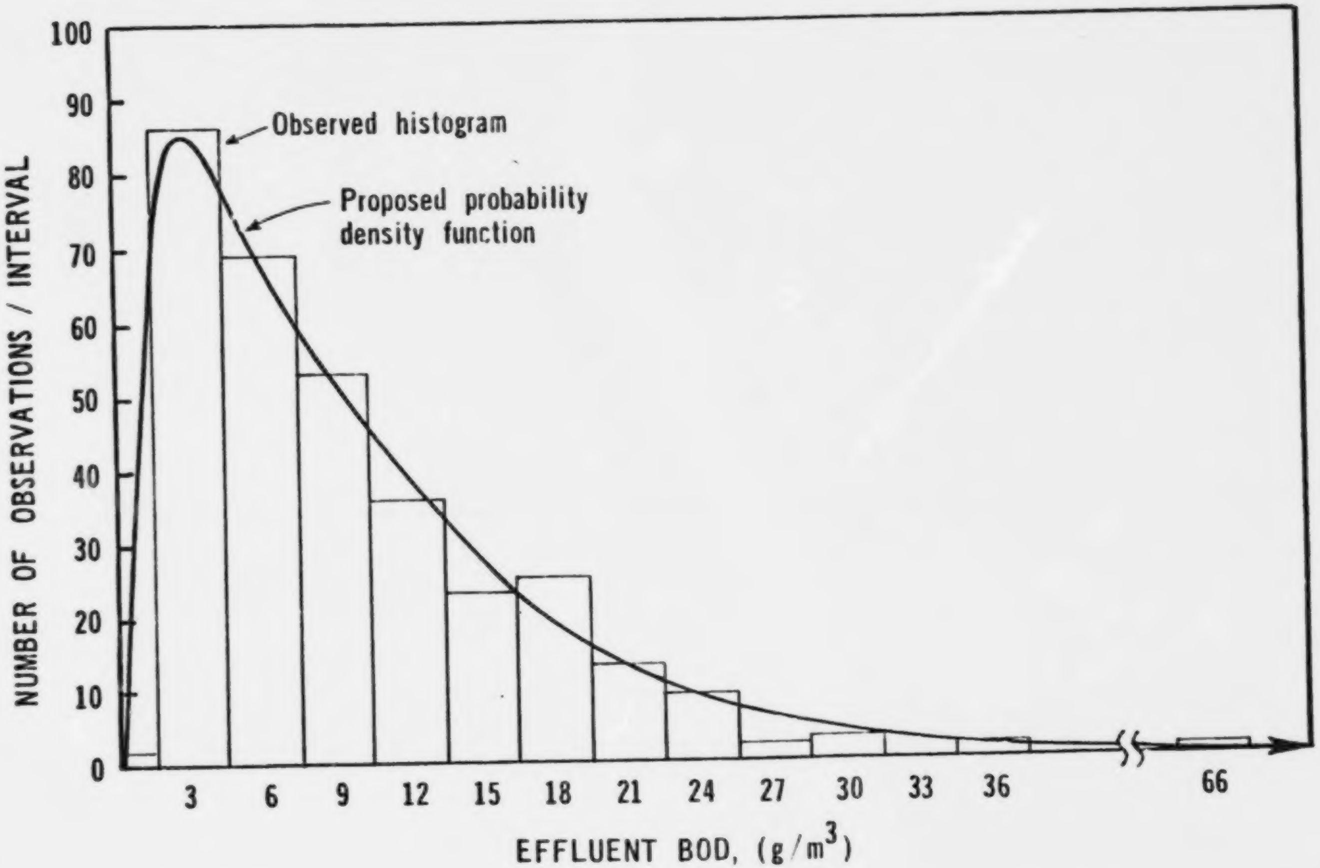


Figure 1. Histogram and probability density function of observed BOD concentration. Plant number 1.

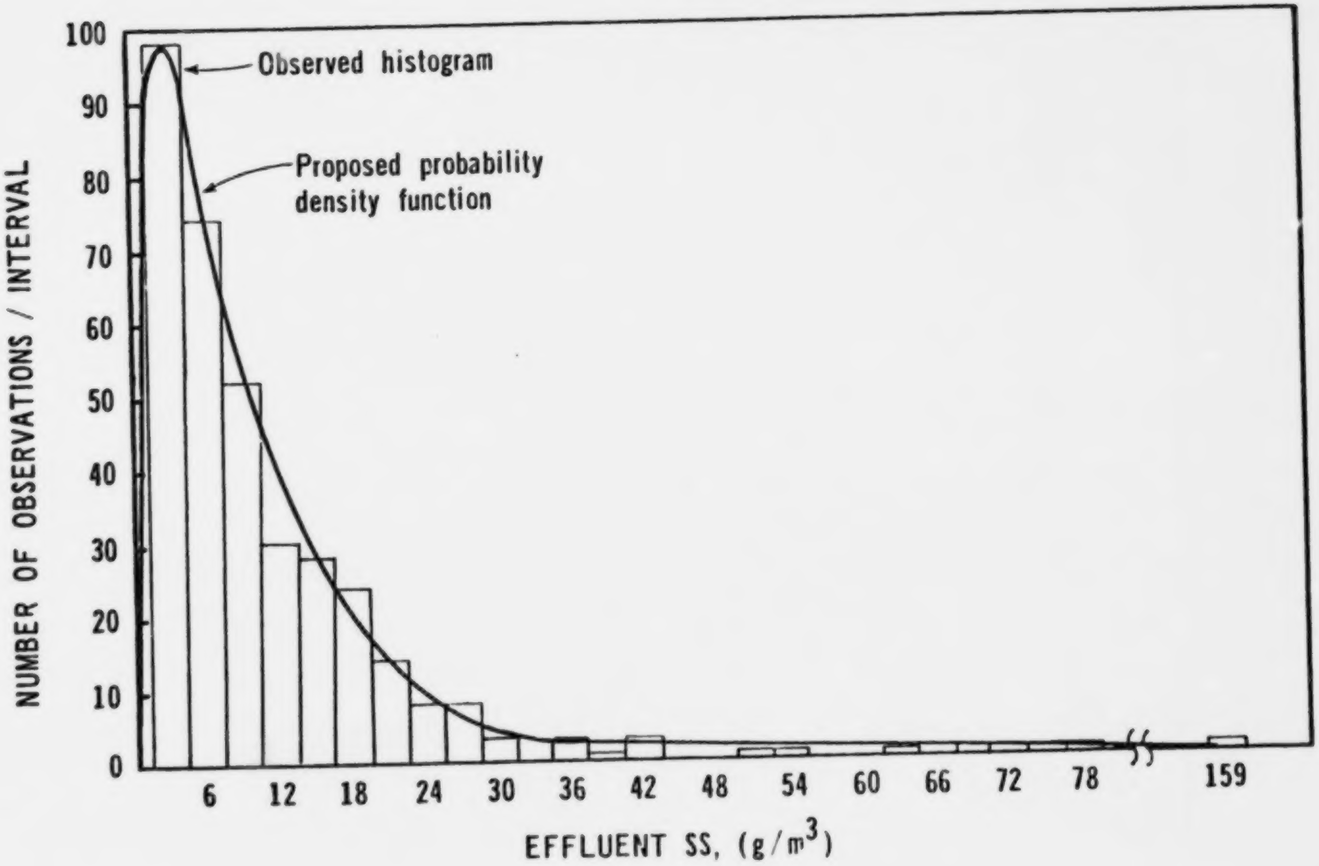


Figure 2. Histogram and probability density function of observed SS concentration. Plant number 1.

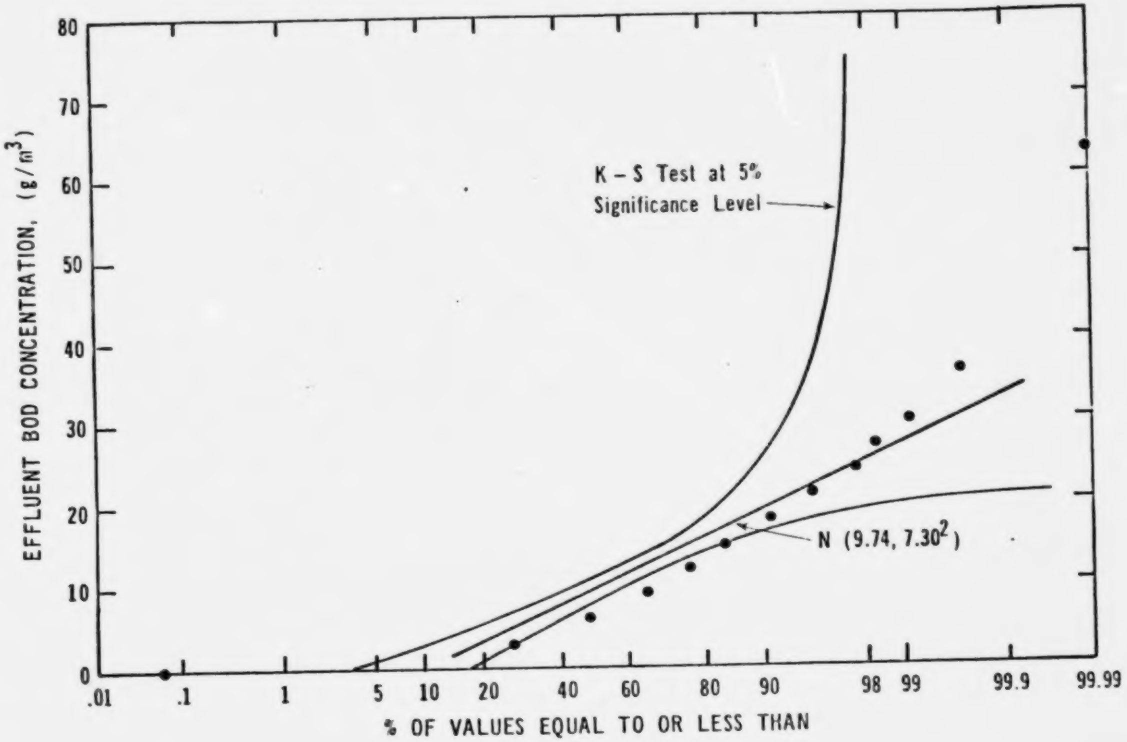


Figure 3. Probability plot for effluent BOD concentration. Plant number 1.

EPA Form 2260-4 (4-80)
TYPING GUIDE SHEET

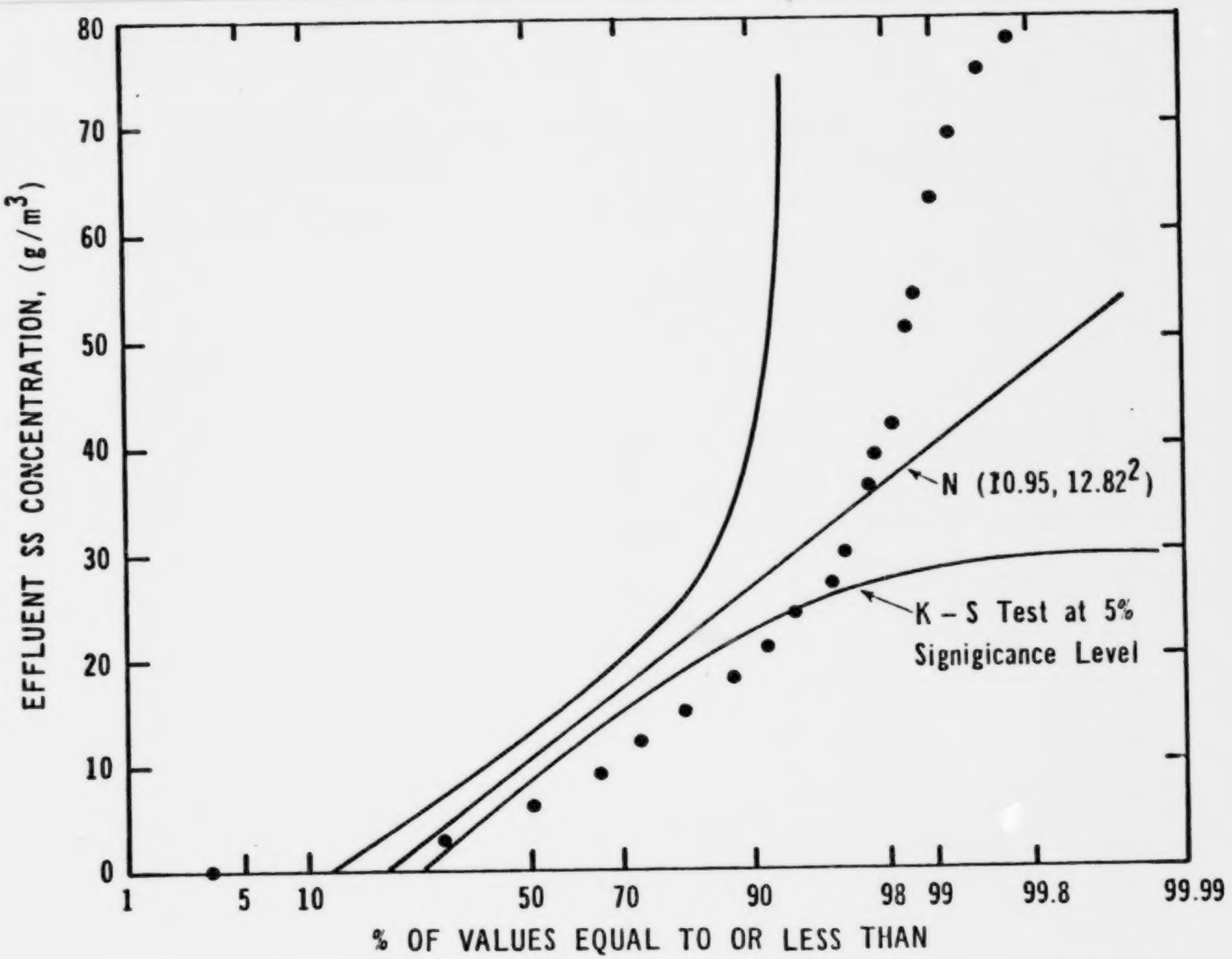


Figure 4. Probability plot for effluent SS concentration. Plant number 1.

Effluent concentration distribution may be examined from the statistics of the data (Table 2). For instance, the original data for all plants are skewed positively which indicates that all distributions are extended further to the right of the most frequent value. There is also a shift away from the origin (min > 0) in most cases. Theoretically, positive skewness (distribution toward higher values) is to be expected since there is usually lower bounds on effluent concentration (no negative values for effluent concentration), but there are no upper limits.

Candidate distribution functions that can be fitted to skewed data include lognormal and gamma distributions. (8) The lognormal probability distribution is one commonly used in civil engineering practice and seems to have been adopted originally to produce a better fit to skewed data by using this simple transformation of the familiar normal distribution. (8) The lognormal distribution has an important justification from the normal distribution, in that fluctuations from the mean are proportional rather than additive. (4,8,13) A detailed description of the lognormal distribution is given in Appendix C.

Use of the lognormal distribution is often motivated by the fact that the underlying random process is multiplicative in nature. The variables observed here, effluent BOD for example, are affected by many factors, some of which are influent characteristics, some environmental, and others of which are related to the conditions of operation within the plant. When an observed variable may be viewed as the product of many similarly distributed variables, the central limit theorem may be used to infer that the sum of the logarithm of these variables will be approximately normally distributed. (4,8)

All currently used kinetic models for the activated sludge process result in multiplicative expressions for effluent BOD (16,25,26,28,34) rate of BOD removal (e.g., $\alpha = -kCX$ or $-kCX/(K+C)$ where C=organic concentration and X=cell concentration). It is thus not unreasonable to expect effluent BOD to have a lognormal distribution. Moreover, while predictive models for effluent suspended solids do not exist, a multiplicative type relationship would be expected for such parameters as hydraulic and solids loading. It will be shown below that a lognormal distribution is also suitable for describing effluent suspended solids concentrations. The probability plots of the logarithm of effluent BOD and SS for the same plant are shown in Figures 5 and 6. The plots are nearly linear, indicating that a lognormal distribution can be fitted to the effluent BOD and SS data.

Choosing a distribution function (normal, lognormal, Poisson, etc.) should not be arbitrary. Hypothesized models need to be verified using a *goodness of fit test*. Such methods test a *null hypothesis*, that is, the variable is distributed in a particular manner (e.g., lognormal) and has specific parameter values (e.g., mean and standard deviation). The Kolmogorov-Smirnov (K-S) test is commonly used to measure goodness of fit. This test was chosen for this study because a comparison of the hypothesized cumulative distribution to the observed cumulative histogram is used and the observed cumulative histogram was a fundamental part of the data base. It should be emphasized that the Kolmogorov-Smirnov test does not discriminate between distributions, but provides an analysis of whether or not a hypothesized distribution gives a good fit.

Goodness of fit tests are made at significance levels, which may be defined as the probability of rejecting a hypothesis that is, in fact, true, a so-called type I error. Thus if the hypothesis should not be rejected at the 5 percent significance

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

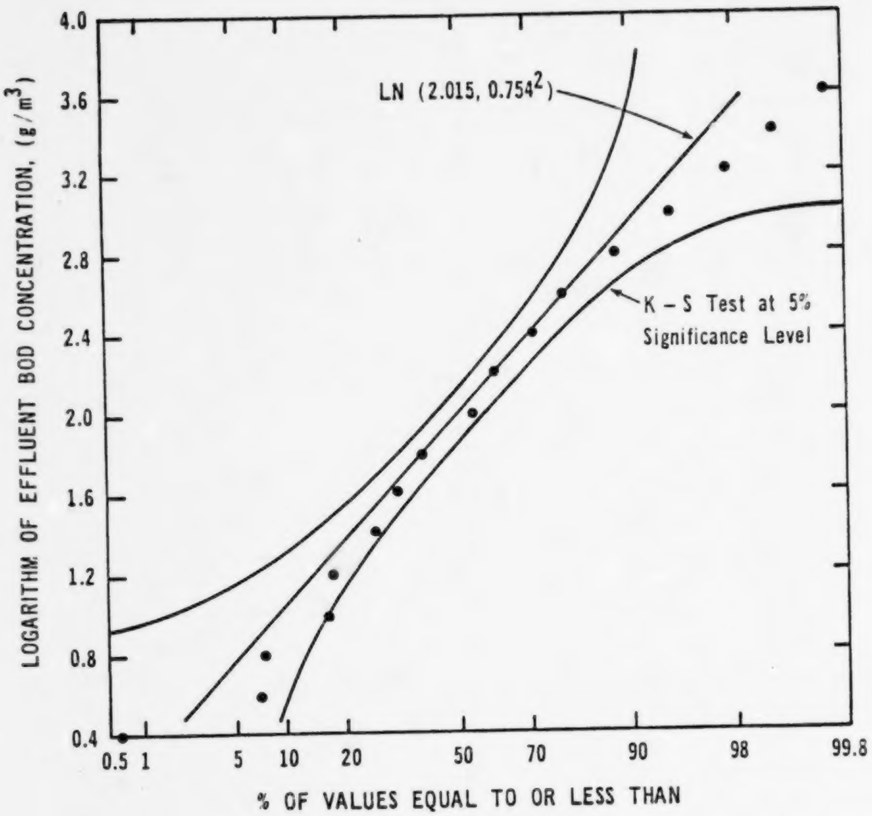


Figure 5. Probability plot for logarithm of effluent BOD concentration. Plant number 1.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
PAGE NUMBER
EPA Form 2350-1 (6-80)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

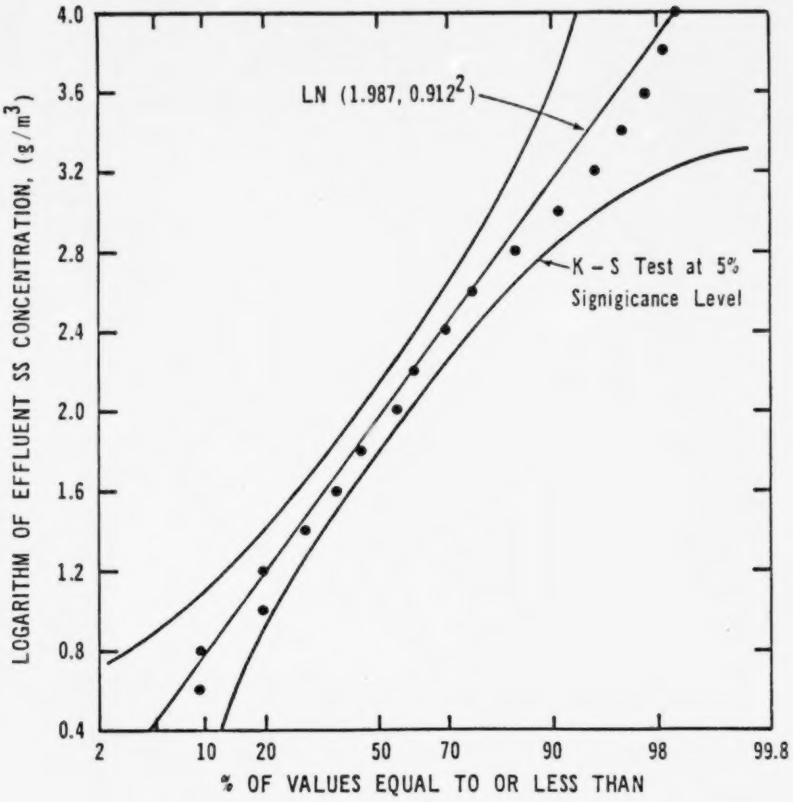


Figure 6. Probability plot for logarithm of effluent SS concentration. Plant number 1.

BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 487)
TYPING GUIDE SHEET

level, it may be assumed that correct hypotheses will be rejected 5 percent of the time (8). Application of the K-S test is described in Appendix D.

The (K-S) test was used to test the goodness-of-fit at the 5 percent level of significance. The graphical execution of the K-S test is shown on all of the probability plots (Figures 3 through 6). In Figures 5 and 6, the hypothesized lognormal distribution is plotted along with curves offset by an amount of $1.36 \sqrt{n}$, the critical value at the 5 percent significance level. (8) Because none of the observed plotting points fall outside this line, the hypothesized lognormal model should not be rejected at the 5 percent significance level. A quantitative goodness-of-fit by K-S test has also resulted in the same conclusion as the graphical one (Table 3). Results of the K-S test for all treatment plants are shown in Table 3 where a one percent significance level was used.

The hypothesized lognormal model should not be rejected for more than two-thirds of the plants at significance levels of one percent. It is important to realize that this significance level is highly acceptable because of the large sample sizes (365) used in this test. As the sample size increases, it is less probable that a set of data pass a certain hypothesized distribution model. (8) Generally, the results reported in Table 3 suggest that a lognormal distribution provides a reasonable description of the effluent BOD and SS data. It should be noted that if for any reason it is assumed that these variables follow the gamma or some other distribution, the conclusion might have been the same; i.e. the distribution is an acceptable model. The purpose of the goodness-of-fit test is to help in accepting or rejecting a given distribution used for prediction. (8) The gamma distribution has also been tested (results are not shown). It was generally found that the lognormal distribution could be used to fit the data of more plants with a higher significance level than the gamma distribution. Popel (33), Garrett (21), and Hann et al (22) have also concluded that effluent BOD and SS concentrations are lognormally (or shifted lognormally) distributed. Berthouex (9), in his discussion of effluent standards, also assumes that effluent concentration data are lognormally distributed.

The importance of determining a distribution is that it can be used to make predictions about future performance with a high degree of confidence, provided that variations appear to come from consistent causes. In practice, occasional large upsets (large deviations from the mean) may occur. These large upsets do not invalidate the distribution function since these upsets are accounted for in a positively skewed (lognormal) distribution.

In Table 3, there are some cases where neither a normal nor a lognormal distribution can be used to describe the data. In such cases, the distribution of effluent concentration should be treated individually. Hovey, et al (23), have developed a linear model based on a simple regression analysis to determine the percentile values for the distributions of effluent BOD and SS concentration exceeding various percentages of the time related to the mean values. Their approach could be used to approximate the distribution of effluent concentration of BOD and SS, when the distribution is not known or does not follow a classical distribution function.

TABLE 3. EFFLUENT BOD AND SS DISTRIBUTIONS KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST[b]

Plant Number	BOD					SS				
	Significance Level For 2-tailed P		Acceptable Distribution[a]			Significance Level For 2-tailed P		Acceptable Distribution[a]		
	Normal	Lognormal	Normal	Lognormal	Other	Normal	Lognormal	Normal	Lognormal	Other
1	0	0.06		X		0	0.14		X	
2	0	0.04		X		0	0			X
3	0	0			X	0	0.03		X	
4	0	0.04		X		0	0.002			X
5	0	0.13		X		0	0.03		X	
6	0	0.25		X		0	0.09		X	
7	0	0			X	0	0.01		X	
8	0	0.01		X		0	0.15		X	
9	0	0.06		X		0	0.05		X	
10	0	0.06		X		0	0.42		X	
11	0	0.04		X		0	0			X
12	0	0.07		X		0	0.23		X	
13	0	0.23		X		0	0.02		X	
14	0	0.06		X		0	0.15		X	
15	0	0.21		X		0	0.32		X	
16	0	0			X	0	0.01		X	
17	0	0			X	0	0			X
18	0	0			X	0	0			X
19	0.03	0.01	X	X		0	0			X
20	0	0.01		X		0	0			X
21	0	0			X	0	0			X
22	0	0			X	0	0			X
23	0	0			X	0	0			X
24	0	0.27		X		0	0.16		X	
25	0	0.44		X		0	0.07		X	
26	0	0.001			X	0	0.04		X	

[a] Indicated distributions not rejected at one percent level of significance

TABLE 3 (continued)

Plant Number	BOD					SS				
	Significance Level For 2-tailed P		Acceptable Distribution			Significance Level For 2-tailed P		Acceptable Distribution		
	Normal	Lognormal	Normal	Lognormal	Other	Normal	Lognormal	Normal	Lognormal	Other
27	0.01	0.01	X	X		0	0.01		X	
28	0	0.03		X		0	0.03		X	
29	0	0			X	0	0.001			X
30	0	0			X	0	0.07		X	
31	0	0.05		X		0	0.12		X	
32	0.003	0.002			X	0	0.04		X	
33	0	0.56		X		0.01	0.04	X	X	
34	0	0.03		X		0.002	0			X
35	0	0.52		X		0	0.43		X	
36	0	0.35		X		0	0.01		X	
37	0	0			X	0	0.06		X	
38	0	0			X	0	0			X
39	0	0.90		X		0	0.41		X	
40	0	0.03		X		0	0.02		X	
41	0	0.06		X		0	0.31		X	
42	0	0			X	0	0.01		X	

[a] Indicated distributions not rejected at one percent level of significance

SECTION 4

RELIABILITY OF ACTIVATED SLUDGE PROCESSES

Effluent quality regulations to restore the quality of receiving waters have been established by the United States Environmental Protection Agency as mandated by the Federal Water Pollution Control Act Amendments of 1972 (P.L. 92-500). (17) Discharge requirements vary from plant to plant, but in general are at least as strict as secondary treatment guidelines. Current regulations often require very high average treatment efficiencies and also restrict variations from the allowable average concentration. In most cases, penalties have been set for violation of these discharge requirements. These requirements can be defined as constraints or limitations on process reliability.

To produce a high effluent quality and to meet the requirements at minimum cost, design engineers must be able to estimate the expected effluent quality and its variations (i.e., the relationship between the occasional high values and the average values) for a given treatment process. Uncertainties and their significance on process performance can be analyzed systematically using methods of probability. A probabilistic approach for design can be used to provide a consistent basis for the analysis of uncertainty and a theoretical basis for the analysis of performance and reliability.

OBJECTIVE

The objective of the analysis reported in this section was to develop a reliability model and to present a simple graphical or tabulated procedure that can be used to predict process performance of plants under design or currently under operation.

RELIABILITY CONCEPT

System reliability can be defined as "the ability to perform the specified requirements free from failure" or "the probability of adequate performance for at least a specified period of time under specified conditions". (11) A treatment plant is completely reliable if process performance response has no failure (e.g. discharge requirement violations). Failure of a treatment process is realized when the required effluent discharge standards are exceeded, i.e.

$$\text{"Failure"} = \text{effluent concentration} > \text{effluent requirements} \quad (5)$$

TYPING GUIDE SHEET

Because of numerous uncertainties underlying the design and operation of a wastewater treatment plant, there is some risk of failure (discharge requirement violation) that is unavoidable. This risk should be recognized and wastewater treatment plants ought to be designed on the basis of an acceptable measure of risk (of violation).

In a technical definition, the essential concept of reliability is "probability of success" or "probability of adequate performance"; i.e., the percent of the time that effluent concentration meets the requirements.

$$\text{Reliability} = 1 - P(\text{failure}) = 1 - P\,(\text{effluent concentration} > \text{requirements}) \tag{6}$$

The probability of failure is extremely sensitive to the distribution function of effluent concentration. Thus, to develop a reliability model, first the distribution of effluent concentration should be modeled. Assuming the distribution of effluent concentration is known, then, the reliability of the plant can be computed. If an expression can be found to define the fraction of the time that a given concentration has been exceeded in the past, the future performance of the plant can be predicted, provided process variables remain the same. (13)

Reliability figures cannot be interpreted until common criteria of performance judgement are established. (11) Regulatory agencies establish these criteria based on adverse effects that a performance failure may cause on the environment.

In addition to computing the probability of failure the minimal reliability requirements must be determined, that is, the question: how large a probability of failure can be accepted must be answered. The classical approach to this question is to apply statistical-economic decision theory (40). The principle of the method is that the total cost of a treatment plant, which is assumed to be the sum of initial cost of the construction and extra cost of daily operation, and the value of the cost of failure multiplied by the probability of its occurrence, should be minimum. In equation form, the total cost is:

$$\text{Total cost} = \text{initial cost} + \text{operational cost} + \text{cost of failure} \times P(\text{failure}) \tag{7}$$

For example, a treatment process with 99 percent reliability is expected to have between 3 to 4 violations in one year. If a treatment plant is designed to allow no more than one violation per year, its reliability should be 99.7 percent. There should be a significant difference in the initial cost and daily operation of process with different limits of reliability. High expectations with respect to effluent quality may necessitate the need for additional or more sophisticated operators, installation of modern control systems, and in some cases, a requirement of tertiary treatment units and/or increasing the physical size of the process. Any of these modifications to the system will add to the initial cost of construction and add extra operational costs. Correspondingly, lesser expectations in water quality or acceptance of higher probability of violation will result in a lesser need for sophisticated operational procedures and, therefore, less cost. In the latter case, the consequences of violation should be considered; tangible as well as intangible costs associated with each violation and the probability of its occurrences should be evaluated. The final decision is a tradeoff between the cost of adverse effects of discharge violation on the environment and the extra initial and operational costs of a more sophisticated process. The degree of reliability (probability of adequate performance) should be selected in such a way that these costs are minimized.

EFFLUENT QUALITY DISTRIBUTION CHARACTERISTICS

Analysis of data, as discussed in the previous section, has resulted in the conclusion that the lognormal assumption for effluent BOD and SS distributions consistently gives a good overall fit to observed effluent values, and thus, lognormal distribution can be used to predict the effluent performance and the reliability of wastewater treatment plants.

The lognormal distribution has the following probability density function (8) (see Appendix C):

$$f_X(x) = \frac{1}{x\sqrt{2\pi}\,\sigma_{LnX}} \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_{LnX}} \text{Ln} \frac{x}{\breve{m}_X}\right]^2\right\} \qquad x \geq 0 \tag{8}$$

where: x = represents a discrete effluent variable concentration

X = is a lognormally distributed continuous random variable with parameters m_X and σ_{LnX}

σ_{LnX} = standard deviation of the natural logarithm of X

$\breve{m}_X$ = median of X

From the properties of the moment-generating function of the normal distribution, the rth moment about the origin is denoted by $E(X^r)$ is given by (4,8)

$$E(X^r) = (\breve{m}_X)^r \exp(\tfrac{1}{2}\, r^2 \sigma^2_{LnX}) \tag{9}$$

In particular,

$$m_X = \breve{m}_X \exp(\tfrac{1}{2}\, \sigma^2_{LnX}) \tag{10}$$

$$\sigma^2_X = m^2_X\, [\exp(\sigma^2_{LnX}) - 1] \tag{11}$$

where m_X and σ^2_X are mean and variance of original data, respectively. Usually the moments of variable X are available. It follows from Equations 10 and 11 that the relationships for the parameters of the probability density function of lognormal distribution in terms of the moments of variable X are (8)

$$\sigma^2_{LnX} = Ln(\frac{\sigma^2_X}{m^2_X} + 1) \qquad (12)$$

and

$$m_{LnX} = Ln\ m_X - \frac{1}{2}\sigma^2_{LnX} \qquad (13)$$

where m_{LnX} = mean of the natural logarithm of X.

Equations 12 and 13 will be used to develop process reliability discussed in the following section.

PROCESS RELIABILITY

Reliability of a treatment plant is based upon knowledge of process behavior. Because of variations in performance, a treatment plant should be designed to produce an average effluent concentration below the discharge standards. The question is, what mean value should be used to guarantee an effluent concentration consistently less than a standard with a certain reliability? A coefficient of reliability (COR) may be introduced that relates mean constituent values (i.e. the design values) to the standard that must be achieved on a probability basis. For example, if a 30 g/m^3 standard is to be achieved 90 percent of the time, the process must be designed to achieve a particular mean effluent concentration. The mean constituent value, m_X, is then obtained from Equation 14

$$m_X = (COR)\ X_s \qquad (14)$$

where X_s = a fixed standard.

Suppose that, for some probability of failure α between zero and one, but presumably near zero, a process is to be designed for which the observed lognormal variable X has the property that

$$P(X \leq X_s) = 1 - \alpha \qquad (15)$$

That is the probability, P, that the standard, X, will be met is $1-\alpha$.

Because the variable X is lognormally distributed, a transformation must be made and this is equivalent to choosing the parameters of the lognormal distribution such that:

$$P\ (Z \leq \frac{\mathrm{Ln}X_s - m_{\mathrm{Ln}X}}{\sigma_{\mathrm{Ln}X}}) = 1 - \alpha \tag{16}$$

where Z is normally distributed with a mean = 0 and a variance = 1.0.

$\frac{\mathrm{Ln}X_s - m_{\mathrm{Ln}X}}{\sigma_{\mathrm{Ln}X}}$ is the standard normal variate, can be obtained from Tables (27) and will be denoted here as $Z_{1-\alpha}$ such that

$$P\ \ (Z \leq Z_{1-\alpha}) = 1 - \alpha \tag{17}$$

Selected values of $Z_{1-\alpha}$ are given in Table 4. Application of Equation (17) is made by selecting a probability value, 1-α. For example, if $1-\alpha = 0.95$ is selected, $Z_{1-\alpha} = 1.645$. The probability that LnX is less than $m_{\mathrm{Ln}X} + 1.645\ \sigma_{\mathrm{Ln}X}$ is then 95 percent.

The problem of interest is to develop an expression for COR that can be used to compute the necessary mean value of the process that guarantees the reliability level 1-α, given a known standard X_s and a known coefficient of variation. Substituting the properties of lognormal distribution (Equations 12 and 13) in Equation 16 results in the following:

$$Z_{1-\alpha} = \frac{\mathrm{Ln}X_s - \left(\mathrm{Ln}\ m_X - \frac{1}{2}\ \mathrm{Ln}(V_X^2 + 1)\right)}{\left(\mathrm{Ln}(V_X^2 + 1)\right)^{\frac{1}{2}}} \tag{18}$$

where V_X is the coefficient of variation. After some algebric manipulations,

$$m_X = (V_X^2 + 1)^{\frac{1}{2}}\ \exp\left\{- Z_{1-\alpha}\left[\mathrm{Ln}(V_X^2 + 1)\right]^{\frac{1}{2}}\right\}(X_s) \tag{19}$$

where

$$= (V_X^2 + 1)^{\frac{1}{2}}\ \exp\left\{- Z_{1-\alpha}\left[\mathrm{Ln}(V_X^2 + 1)\right]^{\frac{1}{2}}\right\} \tag{20}$$

The COR (Equation 20) relates the mean constituent value, m_X, to the standard X_s for reliability level of 1-α. Note that COR is expressed based on properties of the original data and not the logarithm of the data.

TABLE 4. CUMULATIVE PROBABILITIES OF THE STANDARD NORMAL DISTRIBUTION

Cumulative Probability 1-α = Reliability	$Z_{1-\alpha}$
99.9	3.090
99	2.326
98	2.054
95	1.645
92	1.405
90	1.282
80	0.842
70	0.525
60	0.253
50	0

27

TYPING GUIDE SHEET

RELIABILITY BASED DESIGN OF ACTIVATED SLUDGE PROCESSES

Values of the COR for effluent concentrations for different coefficients of variation (V_X) at different levels of reliability are given in Table 5. The results shown in Table 5 have a significant potential application in design. Values of COR given in Table 5 may be used in design of a treatment plant expected to perform at a certain reliability level. The coefficient of variation for effluent concentration should be estimated. Selection of the appropriate design value of the coefficient of variation, V_X, must be based on reasonable expectation of performance (i.e., experience) just as are other design parameters now in use (e.g. hydraulic residence time and overflow rate). This may be done from past experience or using the results from other similar treatment plants. Values of coefficient of variation for effluent BOD and SS for some treatment plants are given in Table 2[a]. Implications derived from Table 5 are best described by an example:

> To have 95% reliability that the effluent BOD concentration will be equal to or less than a certain standard X_s, when V_X for the plant is estimated to be 0.70, the plant should be designed for the mean value equal to or less than 0.43 X_s. (If $X_s = 30$ g/m^3; then average BOD = 0.43 x 30 = 12.9 g/m^3 or less. Similarly, for a less stringent standard of $X_s = 40$ g/m^3; average BOD = 0.43 x 40 = 17.2 g/m^3 or less).

As would be expected, the mean value of required effluent concentration varies inversely with reliability level and with the coefficient of variation in effluent concentration.

Estimation of Reliability of Activated Sludge Processes

The relationships in Table 5 can be viewed in a slightly different context by referring Table 6. Here the fractional reliability for effluent concentration is presented as a function of V_X and the mean to standard concentration ratio (note that this ratio is equal to the COR value). Results from Table 6 may be used to estimate the reliability of a treatment plant under operation if the mean and standard deviation for the plant is known. For example, the reliability of a plant operating with a mean value of 0.6 X_s and a $V_X = 0.70$[a] is 87 percent. In other words, in a long-run operation, about 13 percent of the observed effluent variables would exceed 30 g/m^3 in a plant operating with mean value of 18 g/m^3 and standard deviation of 12.6 g/m^3 ($V_X = 0.70$). Table 6 was constructed based on percentiles $Z_{1-\alpha}$

---

[a] The reader may also consult the results shown in Table 16 in Section 4 of this paper.

TABLE 5. COEFFICIENT OF RELIABILITY (COR) AS A FUNCTION OF V_X AND RELIABILITY BASED ON LOGNORMAL DISTRIBUTION

Reliability (1-α)	50%	80%	90%	92%	95%	98%	99%	99.9%
V_X								
0.3	1.04	0.81	0.71	0.69	0.64	0.57	0.53	0.42
0.4	1.08	0.78	0.66	0.63	0.57	0.49	0.44	0.33
0.5	1.12	0.75	0.61	0.58	0.51	0.42	0.37	0.26
0.6	1.17	0.73	0.57	0.54	0.47	0.37	0.32	0.21
0.7	1.22	0.72	0.54	0.50	0.43	0.33	0.28	0.17
0.8	1.28	0.71	0.52	0.48	0.40	0.30	0.25	0.15
0.9	1.35	0.70	0.50	0.46	0.38	0.28	0.22	0.12
1.0	1.41	0.70	0.49	0.44	0.36	0.26	0.20	0.11
1.2	1.56	0.70	0.46	0.41	0.33	0.22	0.17	0.08
1.5	1.80	0.72	0.45	0.39	0.30	0.19	0.14	0.06

$$^{a}\mathrm{COR} = \frac{m_X}{X_s} = (V_X{}^2 + 1)^{1/2} \exp\left\{-Z_{1-\alpha}\left[\mathrm{Ln}(V_X{}^2 + 1)\right]^{1/2}\right\}$$

example:

$$\left.\begin{array}{l}\text{At } V_X = 1.5 \\ \text{reliability} = 80\%\end{array}\right\} \quad \mathrm{COR} = (1.5^2 + 1)^{1/2} \exp\left\{-.842\left(\ln\,(1.5^2 + 1)\right)^{1/2}\right\} = 1.8028\ e^{-.9141} = \frac{1.8028}{2.4946} = .722$$

TABLE 6. RELIABILITY AS A FUNCTION OF NORMALIZED MEAN AND V_X OF EFFLUENT CONCENTRATION

$\frac{m_X}{X_s}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.5	2.0
V_X												
0.3	>.9999	>.9999	>.9999	0.9995	0.9939	0.9706	0.9134	0.8179	0.6932	0.5583	0.1083	0.0136
0.4	>.9999	>.9999	0.9995	0.9950	0.9767	0.9353	0.8681	0.7799	0.6793	0.5763	0.1950	0.0542
0.5	>.9999	>.9999	0.9973	0.9852	0.9557	0.9062	0.8391	0.7607	0.6770	0.5934	0.2669	0.1093
0.6	>.9999	0.9993	0.9928	0.9732	0.9366	0.8846	0.8212	0.7516	0.6798	0.6092	0.3251	0.1660
0.7	>.9999	0.9979	0.9869	0.9610	0.9212	0.8675	0.8107	0.7483	0.6853	0.6239	0.3721	0.2171
0.8	>.9999	0.9939	0.9805	0.9509	0.9087	0.8594	0.8047	0.7482	0.6920	0.6373	0.4110	0.2632
0.9	.9996	0.9932	0.9743	0.9423	0.9006	0.8527	0.8018	0.7539	0.6992	0.6499	0.4439	0.3033
1.0	.9993	0.9905	0.9688	0.9353	0.8941	0.8485	0.8009	0.7530	0.7065	0.6614	0.4718	0.3398
1.2	.9982	0.9851	0.9597	0.9254	0.8860	0.8445	0.8023	0.7604	0.7201	0.6816	0.4828	0.3967
1.5	.9961	0.9785	0.9507	0.9172	0.8812	0.8446	0.7754	0.7728	0.7389	0.7075	0.5675	0.4620

Example:

At $V_X = 1.0$

$\frac{m_X}{X_s} = 0.5$

$$Z_{1-\alpha} = \frac{-\ln\left(\frac{m_X}{X_s}(V_X{}^2 + 1)^{-1/2}\right)}{[\mathrm{Ln}\,(V_X{}^2 + 1)]^{1/2}} = \frac{-\ln\,(.5\,(2)^{1/2})}{(\ln(2))^{1/2}} = \frac{-\ln\,(.3535)}{(.6931)^{1/2}} = \frac{1.04}{.8325} = 1.249$$

From Tables 27 at Z = 1.249

F(Z) = .8941

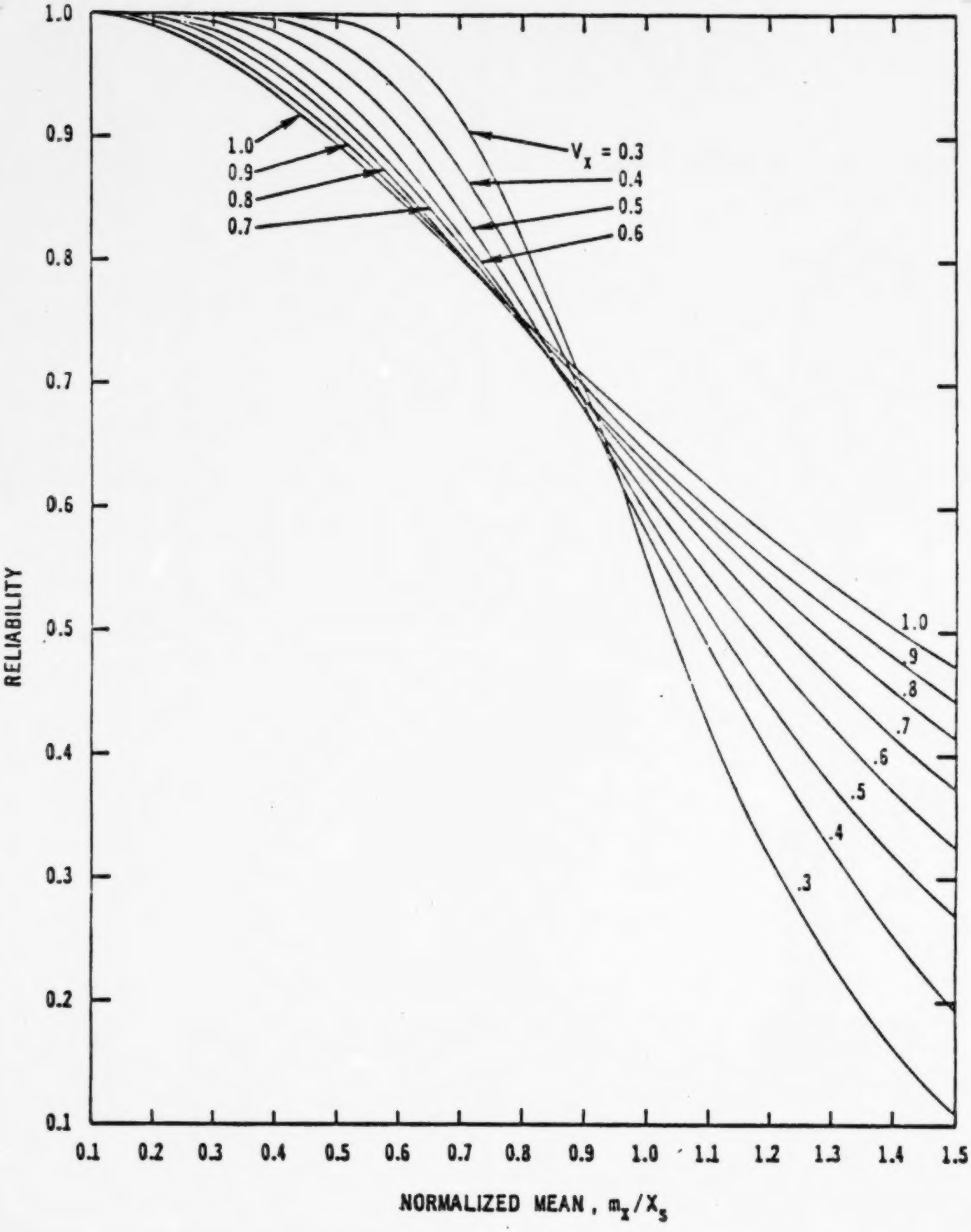


Figure 7. Reliability versus normalized mean for different coefficients of variations.

TABLE 7. PERCENTILE VALUES OF EFFLUENT BOD CONCENTRATION

Plant Number	Effluent BOD Percentiles g/m^3						Normalized Percentiles					
	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile
1	8	15	19	23	30	65	-0.24	0.72	1.27	1.82	2.78	7.57
2	8	12	14	17	21	25	-0.10	0.81	1.27	1.95	2.85	3.76
3	8	13	18	25	37	43	-0.30	0.45	1.20	2.25	4.05	4.95
4	15	24	32	42	57	76	-0.29	0.50	1.20	2.08	3.39	5.06
5	13	22	30	34	46	63	-0.35	0.62	1.49	1.92	3.21	5.05
6	14	24	30	35	41	60	-0.27	0.80	1.45	1.99	2.64	4.68
7	18	26	32	40	72	92	-0.28	0.41	0.94	1.63	4.41	6.15
8	12	21	29	39	56	202	-0.27	0.29	0.77	1.41	2.46	11.53
9	20	38	52	65	94	174	-0.21	0.59	1.21	1.79	3.09	6.65
10	29	37	45	52	65	91	-0.12	0.57	1.26	1.86	2.98	5.22
11	14	22	30	46	97	110	-0.27	0.21	0.70	1.66	4.73	5.51
12	15	20	25	30	42	52	-0.22	0.49	1.20	1.91	3.61	5.03
13	11	25	34	46	62	94	-0.38	0.61	1.25	2.09	3.22	5.49
14	16	32	43	60	93	242	-0.32	0.42	0.92	1.70	3.21	10.03
15	18	42	53	70	105	170	-0.37	0.67	1.15	1.89	3.42	6.26
16	5	6	8	10	14	16	0.04	0.44	1.24	2.03	3.63	4.42
17	1	3	4	5	7	7	-0.68	0.76	1.48	2.20	3.64	3.64
18	2	3	3	4	4	4	-0.35	0.94	0.94	2.22	2.22	2.22
19	5	6	7	9	10	11	0.09	0.59	1.08	2.07	2.57	3.06
20	7	11	13	15	18	20	-0.19	0.87	1.40	1.93	2.73	3.26
21	4	6	7	8	14	18	-0.10	0.74	1.16	1.58	4.10	5.78
22	4	6	7	8	11	14	-0.13	0.86	1.35	1.84	3.32	4.79
23	3	4	5	7	9	10	-0.02	0.60	1.19	2.35	3.53	4.11
24	15	41	60	90	112	140	-0.35	0.62	1.32	2.43	3.25	4.28
25	10	18	24	30	47	61	-0.32	0.57	1.23	1.90	3.78	5.33
26	9	12	15	18	25	28	-0.01	0.67	1.35	2.04	3.63	4.31
27	12	17	21	24	32	38	-0.09	0.68	1.30	1.77	3.00	3.93
28	13	19	23	25	29	48	-0.12	0.79	1.40	1.71	2.31	5.20

EPA Form 2300-4 (4-80)
TYPING GUIDE SHEET

TABLE 7 (continued)

Plant Number	Effluent BOD Percentiles g/m^3						Normalized Percentiles					
	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile
29	2	7	10	11	16	24	-0.60	0.72	1.50	1.77	3.08	5.19
30	4	9	14	19	24	26	-0.35	0.53	1.41	2.29	3.17	3.52
31	9	13	17	29	48	50	-0.22	0.26	0.73	2.14	4.38	4.62
32	23	33	37	42	46	50	-0.10	0.91	1.32	1.83	2.23	2.63
33	13	28	38	51	60	98	-0.31	0.62	1.24	2.04	2.59	4.94
34	8	13	16	19	24	40	-0.15	0.81	1.38	1.95	2.90	5.96
35	25	44	60	70	155	248	-0.24	0.46	1.04	1.41	4.55	7.97
36	5	11	14	17	20	21	-0.37	0.91	1.56	2.20	2.84	3.05
37	20	30	37	44	55	120	-0.08	0.70	0.25	1.80	2.66	7.75
38	9	22	39	55	79	164	-0.38	0.28	1.13	1.94	3.15	7.43
39	70	125	175	200	265	345	-0.27	0.68	1.54	1.97	3.10	4.48
40	13	18	22	26	31	32	-0.27	0.68	1.43	2.18	3.12	3.31
41	8	12	14	16	22	23	-0.27	0.77	1.29	1.82	3.38	3.64
42	15	27	34	45	81	110	-0.40	0.45	0.95	1.73	4.29	6.35
43	13	25	41	63	140	260	-0.31	0.13	0.72	1.53	4.36	8.78
						AVERAGE	-0.24	0.61	1.19	1.92	3.29	5.27

TABLE 8. PERCENTILE VALUES OF EFFLUENT SS CONCENTRATION

Plant Number	Effluent SS Percentiles g/m^3						Normalized Percentiles					
	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile
1	7	16	21	28	62	150	-0.31	0.39	0.78	1.33	3.98	10.84
2	6	8	9	11	14	30	-0.10	0.65	1.02	1.76	2.88	8.82
3	14	18	20	23	26	28	-0.12	0.75	1.19	1.84	2.49	2.93
4	19	36	50	64	113	158	-0.32	0.45	1.09	1.73	3.97	6.03
5	17	26	36	46	86	145	-0.27	0.31	0.95	1.58	4.13	7.90
6	18	28	34	45	98	139	-0.24	0.38	0.75	1.43	4.72	7.26
7	11	21	32	40	110	213	-0.26	0.20	0.72	1.10	4.40	9.24
8	15	25	36	65	123	224	-0.23	0.15	0.57	1.66	3.84	7.65
9	24	76	104	140	220	332	-0.42	0.65	1.22	1.96	3.61	5.70
10	56	71	80	93	114	150	-0.15	0.65	1.13	1.83	2.95	4.87
11	6	11	17	36	78	186	-0.25	0.04	0.38	1.46	3.86	10.01
12	12	20	26	31	54	136	-0.23	0.46	0.98	1.41	3.40	10.50
13	10	21	35	46	75	143	-0.28	0.37	1.21	1.86	3.60	7.67
14	20	45	64	119	211	534	-0.31	0.21	0.61	1.76	3.69	10.46
15	15	38	74	96	160	244	-0.38	0.29	1.34	1.99	3.87	6.34
16	5	7	9	11	13	23	-0.12	0.56	1.25	1.93	2.61	6.02
17	1	3	7	9	13	22	-0.54	1.27	1.54	2.24	3.63	6.75
18	4	6	8	8	11	12	-0.23	0.71	1.65	1.65	3.06	3.52
19	8	11	13	16	19	24	-0.05	0.71	1.22	1.98	2.75	4.02
20	6	8	10	13	23	56	-0.22	0.22	0.67	1.34	3.58	10.96
21	6	8	10	12	18	40	-0.09	0.24	1.00	1.54	3.18	9.17
22	5	7	9	11	17	27	-0.25	0.39	1.04	1.69	3.63	6.87
23	3	5	6	7	10	15	-0.30	0.69	1.18	1.67	3.15	5.62
24	10	28	45	63	95	220	-0.35	0.36	1.04	1.76	3.03	8.01
25	10	16	20	26	54	151	-0.15	0.32	0.64	1.11	3.31	10.95
26	8	14	17	22	28	54	-0.26	0.68	1.15	1.93	2.87	6.95
27	8	12	17	20	26	30	-0.16	0.56	1.47	2.01	3.10	3.83
28	7	13	16	21	38	62	-0.27	0.56	0.98	1.67	4.03	7.36

TABLE 8 (continued)

Plant Number	Effluent SS Percentiles g/m^3						Normalized Percentiles					
	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile	50%ile	80%ile	90%ile	95%ile	99%ile	100%ile
29	9	14	17	20	29	37	-0.17	0.69	1.21	1.73	3.28	4.66
30	8	15	19	23	30	55	-0.29	0.70	1.26	1.83	2.81	6.34
31	10	14	16	19	24	48	-0.19	0.65	1.06	1.69	2.73	7.74
32	22	30	38	43	52	62	-0.23	0.63	1.50	2.04	3.01	4.09
33	18	30	38	48	78	114	-0.23	0.50	0.99	1.60	3.43	5.63
34	14	20	24	27	38	73	-0.10	0.68	1.20	1.60	3.01	7.57
35	21	56	86	118	177	264	-0.38	0.54	1.33	2.18	3.73	6.03
36	6	10	12	14	22	27	-0.33	0.70	1.22	1.74	3.80	5.10
37	14	35	41	47	58	61	-0.34	1.08	1.49	1.91	2.66	2.87
38	30	94	168	212	408	780	-0.39	0.33	1.16	1.65	3.86	8.05
39	54	116	172	222	360	492	-0.35	0.48	1.23	1.91	3.76	5.53
40	7	12	16	20	27	34	-0.34	0.60	1.36	2.12	3.44	4.77
41	8	14	18	24	34	39	-0.27	0.62	1.21	2.09	3.56	4.30
42	17	32	62	81	206	311	-0.31	0.10	0.91	1.42	4.79	7.63
43	21	40	89	199	368	605	-0.30	-0.03	0.66	2.21	4.60	7.94
						AVERAGE	-0.26	0.50	1.08	1.75	3.48	6.85

$$Z_{1-\alpha} = - \frac{\mathrm{Ln}\left[\frac{m_X}{X_S} \; (V_X^2 + 1)^{-\frac{1}{2}}\right]}{\left[\mathrm{Ln}(V_X^2 + 1)\right]^{\frac{1}{2}}} \qquad (21)$$

and reliability has been determined using the standard normal table (e.g., Table 4).

Graphical Form of the Reliability Model

The results of Tables 5 and 6 are presented graphically in Figure 7. This figure can be used in both estimation of the mean value in design of a process or prediction of effluent quality in a plant under operation.

VERIFICATION OF THE RESULTS

To determine the validity of the theoretical results of the COR, Table 5 was reconstructed based on the one year of operation data of 43 treatment plants. Percentile values of exceedance the effluent BOD or SS concentration greater than x percent of the data values for effluent BOD and SS from one year of daily operation data were then computed to present an empirical distribution of effluent concentration without considering whether each set of data follows any classical distribution function. Effluent BOD and SS percentiles for different plants are given in Tables 7 and 8 respectively. The percentile values were also normalized by deducting the mean from each value and dividing by the corresponding standard deviation. For example, the normalized effluent BOD 100%ile value for plant number 1 is 7.57. This value is calculated by deducting the mean effluent BOD (9.74 g/m^3) and dividing by the standard deviation (7.30 g/m^3), both of which were given in Table 2. The average of the normalized values at each percentile was computed and plotted against percentile values on Figure 8. The relationship between reliability versus normalized values based on actual data are shown in this figure. Results of Tables 7 and 8 and Figure 8 were used in recalculation of COR based on data.

TABLE 9. COEFFICIENT OF RELIABILITY[a] AS A FUNCTION OF V_X AND RELIABILITY FROM POOLED PLANT DATA

Row 1 for BOD, Row 2 for SS

Reliability	50%	80%	90%	95%	99%	100%
V_X						
0.3	1.08 1.08	0.85 0.87	0.74 0.75	0.64 0.66	0.50 0.49	0.39 0.34
0.4	1.11 1.11	0.80 0.84	0.68 0.70	0.57 0.59	0.43 0.42	0.32 0.27
0.5	1.14 1.15	0.77 0.80	0.63 0.65	0.51 0.53	0.38 0.36	0.27 0.23
0.6	1.17 1.18	0.73 0.77	0.58 0.61	0.46 0.49	0.34 0.32	0.24 0.20
0.7	1.20 1.22	0.70 0.74	0.55 0.57	0.43 0.45	0.30 0.29	0.21 0.17
0.8	1.24 1.26	0.67 0.72	0.51 0.54	0.39 0.42	0.28 0.26	0.19 0.15
0.9	1.28 1.30	0.65 0.69	0.48 0.51	0.37 0.39	0.25 0.24	0.17 0.14
1.0	1.32 1.35	0.62 0.67	0.46 0.48	0.34 0.36	0.23 0.22	0.16 0.13

[a] $$\text{COR} = \frac{\bar{X}}{X_s} = \frac{1}{Z_{1-\alpha} \cdot V_X + 1}$$

Example Calculation: For 80% Reliability and $V_X = 0.5$

From Table 7, Average $Z_{1-\alpha}$ for BOD data is 0.01

$$\text{COR} = \frac{1}{0.61\ (0.5) + 1} = 0.77$$

TYPING GUIDE SHEET

TABLE 10. COMPARISON OF PREDICTED AND MEASURED PERFORMANCE OF ACTIVATED SLUDGE PLANTS ACROSS THE UNITED STATES

	BOD				SS			
			%time < 30 g/m^3				%time < 30 g/m^3	
Plant Number	V_X	$\bar{X}/30$	Predicted	Measured	V_X	$\bar{X}/30$	Predicted	Measured
1	0.75	0.33	97.10	99.1	1.17	0.37	93.80	96.2
2	0.52	0.28	99.80	100.0	0.43	0.21	>99.99	100.0
3	0.67	0.33	98.30	97.8	0.32	0.48	99.40	100.0
4	0.62	0.61	87.50	87.9	0.83	0.87	71.10	75.5
5	0.57	0.54	92.30	91.2	0.74	0.71	80.20	85.5
6	0.56	0.55	92.60	91.4	0.73	0.73	78.10	83.3
7	0.54	0.71	82.40	86.5	1.27	0.56	86.00	89.9
8	0.99	0.55	87.20	91.5	1.26	0.70	80.40	86.1
9	0.91	0.83	73.50	70.3	0.98	1.48	52.70	58.0
10	0.38	1.01	56.30	60.1	0.32	1.96	2.30	3.1
11	0.90	0.62	84.30	90.4	1.70	0.34	93.40	94.2
12	0.43	0.55	95.10	95.0	0.79	0.49	91.50	94.3
13	0.87	0.55	88.00	88.6	1.13	0.49	89.20	89.0
14	0.95	0.77	76.60	77.9	1.37	1.16	64.40	67.9
15	0.86	0.89	70.20	68.8	1.21	0.94	70.50	70.6
16	0.51	0.16	99.99	100.0	0.55	0.18	>99.99	100.0
17	0.72	0.06	99.99	100.0	1.13	0.09	>99.99	100.0
18	0.34	0.08	99.99	100.0	0.47	0.15	>99.99	100.0
19	0.42	0.16	99.99	100.0	0.48	0.27	>99.99	100.0
20	0.49	0.26	99.90	100.0	0.64	0.23	99.70	99.5
21	0.56	0.14	99.99	100.0	0.58	0.21	99.90	99.5
22	0.48	0.14	99.99	100.0	0.53	0.19	>99.99	100.0
23	0.58	0.10	99.99	100.0	0.56	0.12	>99.99	100.0
24	1.11	0.81	75.30	73.7	1.33	0.63	83.20	81.9
25	0.70	0.43	95.10	95.2	1.07	0.40	93.10	97.0

TABLE 10 (continued)

Plant Number	BOD				SS			
			%time < 30 g/m^3				%time < 30 g/m^3	
	V_X	$\bar{X}/30$	Predicted	Measured	V_X	$\bar{X}/30$	Predicted	Measured
26	0.48	0.30	99.80	100.0	0.66	0.32	98.60	98.9
27	0.51	0.42	97.90	98.3	0.62	0.30	99.20	100.0
28	0.48	0.46	97.30	99.2	0.80	0.30	98.10	98.1
29	0.89	0.14	99.80	100.0	0.58	0.33	99.00	99.5
30	0.95	0.20	99.20	100.0	0.70	0.34	97.90	98.9
31	0.78	0.36	96.60	95.3	0.44	0.36	99.60	98.9
32	0.41	0.80	77.80	72.5	0.38	0.80	78.60	80.0
33	0.90	0.60	85.20	84.2	0.75	0.73	73.10	80.9
34	0.60	0.29	99.40	99.7	0.52	0.49	95.60	97.3
35	0.86	1.05	62.00	64.8	1.07	1.18	59.80	64.9
36	0.70	0.22	99.70	100.0	0.53	0.24	99.90	100.0
37	0.61	0.70	82.00	84.3	0.76	0.64	84.10	76.0
38	1.20	0.55	86.50	86.5	1.37	2.17	40.50	51.1
39	0.68	2.85	8.20	12.1	0.93	2.67	19.80	21.9
40	0.37	0.48	98.70	98.8	0.60	0.29	99.40	99.6
41	0.43	0.30	99.90	100.0	0.69	0.33	98.20	97.7
42	0.68	0.69	81.90	85.8	1.30	0.95	70.80	77.8
43	1.27	0.71	79.90	84.2	1.68	1.41	61.10	71.6

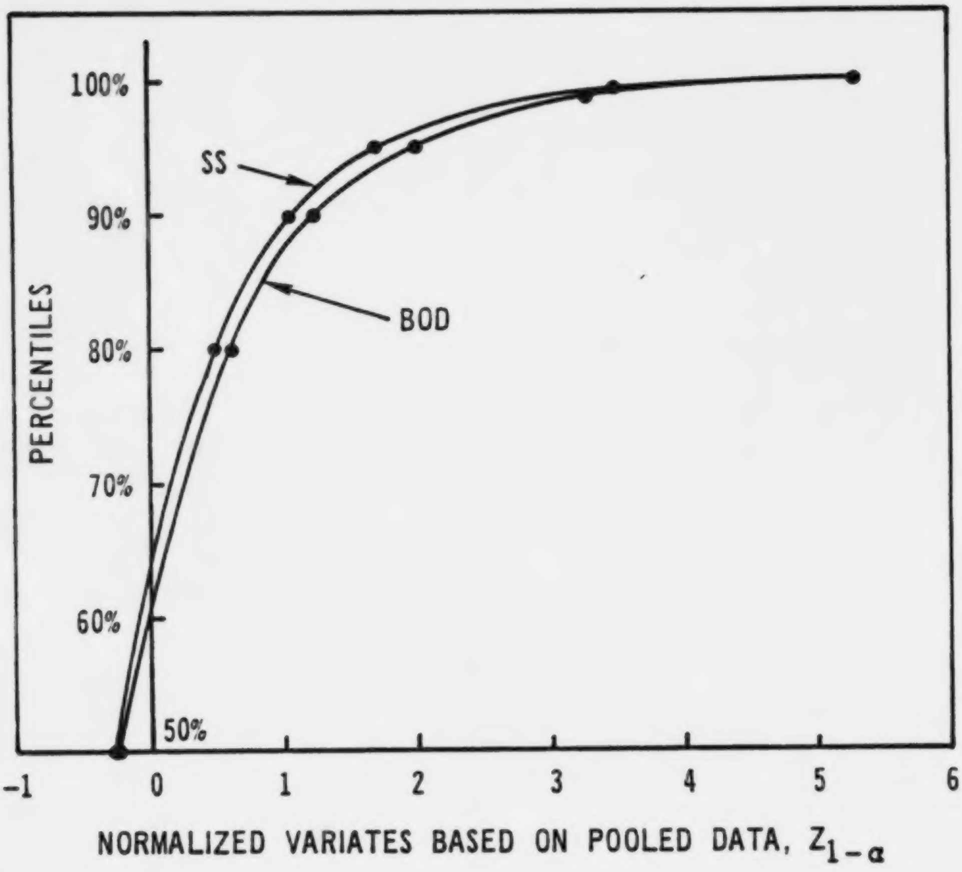


Figure 8. Percentile values versus normalized variates for effluent BOD and SS concentration.

TYPING GUIDE SHEET

The COR based on actual plant data[a] is given in Table 9 for different coefficients of variation of effluent BOD and SS at different levels of reliability. This Table was constructed exactly like Table 5 but on the basis of the results of pooled plant data using Figure 8, rather than lognormal distribution assumption. The results are highly comparable to theoretical results based on a lognormal distribution assumption. In most cases, the results are the same or comparable with a difference less than 5 percent. Usually, the lognormal assumption is conservative, that is the assumption results in a lower reliability prediction than determined from the process data.

To support the validity of the model in prediction of effluent quality, the percent of the time that effluent concentration exceeded from 30 g/m^3 was computed for one year of data in all plants (Table 10). This percent of exceedance was also predicted using the reliability model (Equation 21, Table 6, or Figure 7) based on the lognormal assumption. The predicted values are again very comparable with the measured values.

It should be noted that the reliability predictions are valid only on a long-run operation (e.g., yearly basis), since the effect of nonrandomness, trends and cyclical variations of effluent data have not been considered in the reliability model. To predict the short-term reliability, the long-term trends and non-random components from the data should be removed.

---

[a] The COR based on data is computed as follows:

$$P(X \leq X_s) = P(Z \leq \frac{X_s - \bar{X}}{S_X}) = 1 - \alpha$$

or

$$Z_{1-\alpha} = \frac{X_s - \bar{X}}{S_X}$$

$$\bar{X} = \frac{1}{Z_{1-\alpha} \cdot V_X + 1} X_s$$

and

$$COR_{(data)} = \frac{1}{Z_{1-\alpha} \cdot V_X + 1} \qquad (22)$$

where the percentile $Z_{1-\alpha}$ is calculated from data (Figure 8 and Tables 7 and 8).

SECTION 5

STABILITY OF ACTIVATED SLUDGE PROCESSES

Conventional criteria and design procedures employed in the design of municipal and industrial wastewater processes are based on the simplified assumption that steady-state conditions exist with respect to the influent wastewater characteristics. In practice, activated sludge systems are seldom operated under steady-state conditions. Effluent quality depends on input loads, environmental conditions, in-plant biological and operational variables as well as many other unexpected perturbations beyond operational control. Designing a treatment process is thus a multi-variate problem. Fluctuations in input loads and environmental conditions, nature of the wastewater, introduction of toxic materials, and mechanical failure of the system as well as human errors may frequently cause process upset and instability which may have adverse effects on effluent quality.

PROCESS STABILITY

The term "process stability" has often been used to describe plants with consistent performance without giving a formal definition. Stability may be defined in terms of the magnitude of the perturbations of effluent quality characteristics from the mean values. For example, in a stable process, when fluctuations exist in input loadings, environmental conditions, and operational parameters, large variations will not appear in the effluent quality characteristics.

Instability in the wastewater treatment process may be evidenced by the occurrence of gross failures such as bulking activated sludge or accumulation and overflow of sludge in secondary clarifiers. Reduction in efficiency of the process treatment, which may be followed by a noticeable increase in effluent suspended solids discharged from the final clarifier, is another indication of instability. To meet effluent discharge guidelines, activated sludge process behavior must be examined with respect to variability and thus, stability.

STABILITY VERSUS RELIABILITY

Stability, as defined here, is the extent of variation of effluent characteristics, whereas reliability is defined as the probability of meeting discharge requirements. Therefore, reliability is based on an arbitrary standard and may change if the discharge standards are changed.

Process performance may be defined as stable but not reliable when the extent of variation of effluent quality characteristics is low, whereas discharge requirements are violated (Figure 9-a). A process may also be reliable but not stable when the effluent concentrations have a large standard deviation, but the arbitrary requirements are met (Figure 9-b). In the latter case, there is a high probability that violations will occur, and/or process becomes unbalanced. A process with a low variation in effluent concentration that performs under the upper limit standards is defined to be a stable and reliable process (Figure 9-c). A process with a wide range concentration that sometimes is below and other times above the standard limits is both unstable and unreliable (Figure 9-d).

BACKGROUND

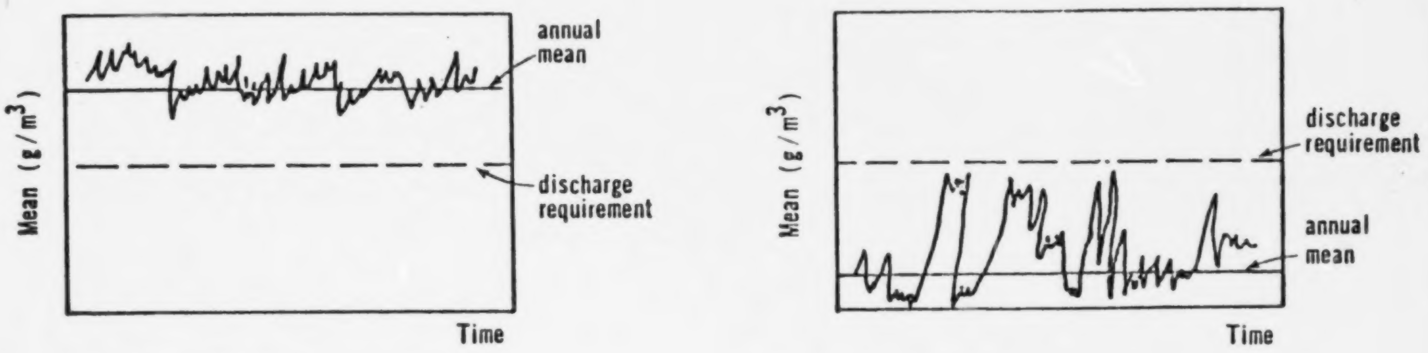
Several approaches are available for the analysis of the process stability: experimental studies under laboratory conditions, mathematical modeling and simulation techniques, or statistical measures. Most of the related studies of stability in the literature have been based on laboratory experiments and mathematical modeling. (5,10,24,36,39)

The effect of shock loads (24) and transient loading (36,39) on effluent quality have been investigated experimentally. Usually these experiments have been done under favorable laboratory conditions permitting investigation of individual variables and thus do not reflect the variations encountered in full-scale operations. Because of these factors, pilot plant or laboratory-scale studies are less likely to replicate the inherent variability and the conditions encountered in the full-scale operations. Therefore, working with full-scale plant data is preferable where the objective is to develop a measure of actual system stability.

Deterministic modeling and/or computer simulation is another technique occasionally used for quantitative description of waste treatment process performance. (5,10) Deterministic models and simulation results should be used with caution in stability studies since these techniques are usually based on kinetic models. Most kinetic models currently used in process design are based on the assumption of steady-state conditions, whereas the concern in stability studies is on non-steady-state conditions. For example, Chiang (10) has developed a stability indicator to examine the process performance stability. Unfortunately, a number of assumptions were made in the development of the non-steady-state model that only hold if steady-state conditions exist; therefore both the model and the results of the study are of limited value. (19)

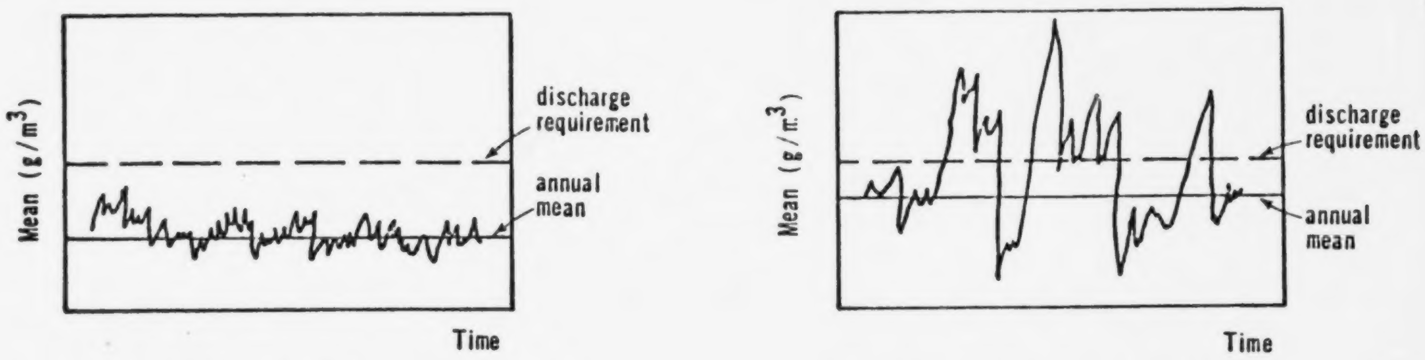
Mathematical and simulation techniques are also subject to the same weaknesses as experimental procedures are; that is, variables that will exist in the full-scale operations may easily be overlooked. For complex systems such as wastewater treatment processes, the fact that a large number of complex differential equations may be necessary to describe the performance and stability of the system often leads to the development of complex models which have proven to be impractical.

Because of the intricate nature of the stability problem and the fact that many of the parameters affecting effluent quality are inherently stochastic, an obvious approach to analyzing process stability is through statistical analysis. Statistical



a) a stable but not reliable process

b) a reliable but not stable process



c) a stable and reliable process

d) an unstable and unreliable process

Figure 9. Reliability versus stability

examination techniques can be used to identify and quantify numerical characteristic existing in the data. Dealing with data from full-scale plants, the stability of a process may be evaluated both qualitatively and quantitatively. In this case, all variables affecting the process variation are considered, regardless of their cause and effect relationships.

OBJECTIVE

The objective of this portion of the study was to evaluate the stability of activated sludge wastewater treatment process based on statistical measures and to introduce a method that can be used to define the stability of a plant and compare the stability among plants.

STABILITY MEASURE

Stability may be thought of as the property of adherence to a reference or norm. In the case of wastewater treatment, the best reference value for stability appears to be the annual mean constituent concentration. One complete cycle of weather and a monitoring period long enough to damp the effects of short term variations is reflected in the annual mean value. Variations from the mean on a day-to-day (or week-to-week, or month-to-month) basis then provide a measure of stability.

Since stability is a measure of variations from the mean, for comparison purposes, a standard quantitative measure of variation (stability indicator) is of special interest. The stability indicator may be used as a tool to evaluate and compare the relative stability of different plants and to examine the effect that various design and operation procedures may have on process stability. Several such indicators are available. Explanation of the significance of each indicator follows:

Variance and Standard Deviation--The most common measures of dispersion, as noted in Section 2, are "variance" and "standard deviation". Standard deviation may be used as a logical indicator for stability comparison of plants. Smaller standard deviations generally imply tighter distributions, less widely spread about mean, and therefore a more stable situation. This has been shown schematically in Figure 10 where probability density functions of the same basic shape but having different means or standard deviations are shown. (8) The curves in Figure 10-a and b differ only in their mean, while the curves in Figures 10-b and c differ only in their standard deviations. The curve in Figure 10-c which has a wide spread corresponds to a less stable situation as compared to the curve shown in Figure 10-b.

Coefficient of Variation--Coefficient of variation is another common measure of variability but is not a good and logical measure of stability. For example, two plants may have the same coefficient of variation, say, 0.7, where one plant has a mean of 50 g/m^3 and a standard deviation of 35 g/m^3, and the other plant has a mean of 10 g/m^3 and a standard deviation of 7 g/m^3. Obviously, the latter plant has less variation and thus, is more stable than the former one.

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

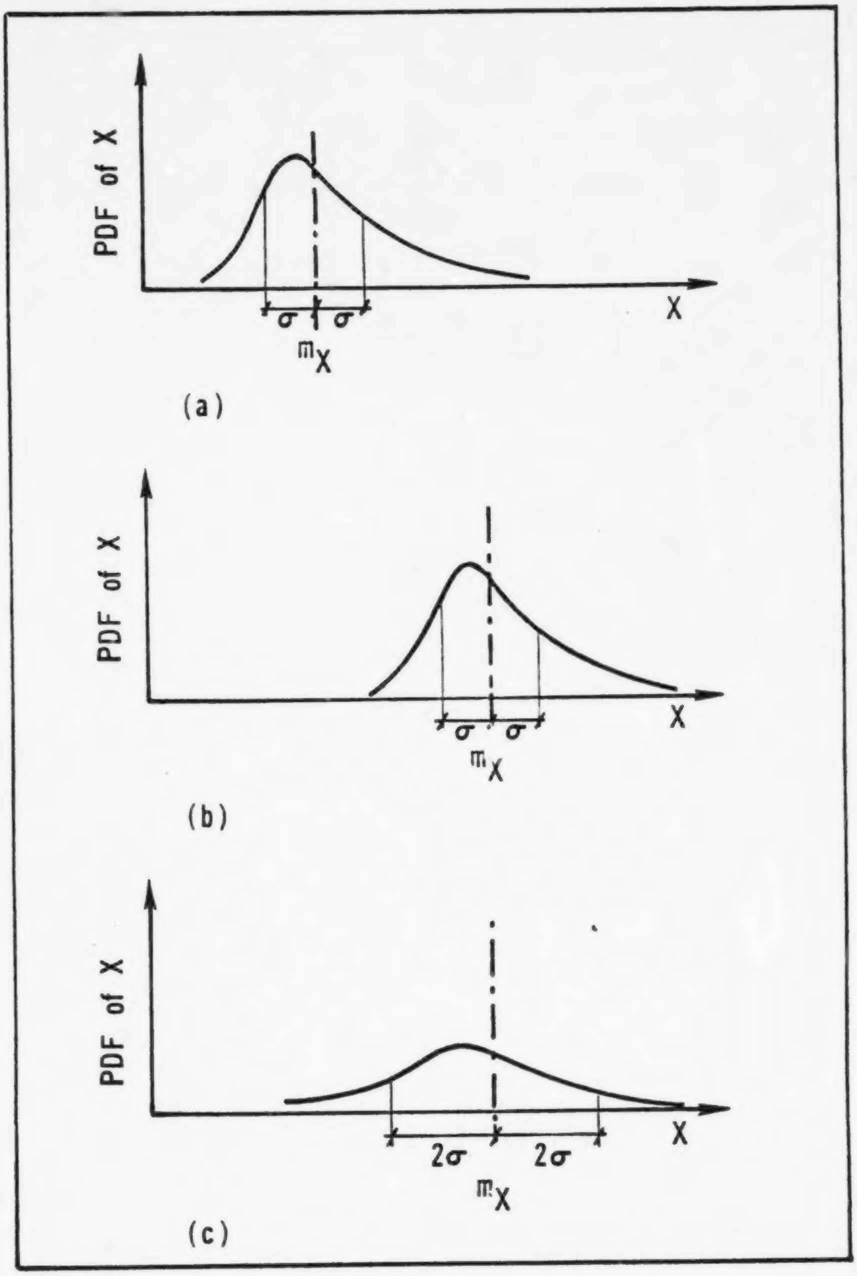


Figure 10. Changes in probability density function (PDF) with changes in means and standard deviations.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2220-4 (4-80)
TYPING GUIDE SHEET

Coefficient of Stability—Forsyth (20) has proposed the "coefficient of stability (COS)," defined as the ratio of variance over the mean, for stability analysis. This coefficient is a far better measure of stability than the coefficient of variation as it gives more weight to the variations while taking the mean into account. In other words, plants that are operating at higher mean concentration are allowed to have more variation than those operating at lower mean concentration with the same measure of stability.

Logarithm of Data—In the previous section, it was concluded that effluent BOD and SS concentrations are lognormally distributed. It is therefore reasonable to suggest that the logarithm of data may be used in measuring stability rather than the data themselves. But this procedure is not an accurate measure. For example, between two plants, if one operates with effluent BOD in the range of 1 to 30 g/m^3 and the other one in the range of 10 to 200 g/m^3, the range of their logarithms will be between 0 to 3.4 and 2.3 to 5.3, respectively. Obviously, the measure of standard deviation for these two ranges will be very close and could be misleading for use in stability measurement. The anti-logarithm of the standard deviation of the logarithm of data [antilog (S_{LnX})] could be a possible measure of stability (14), however.

SELECTION OF THE BEST INDICATOR OF STABILITY

"Standard deviation" has been selected as the most appropriate indicator of wastewater treatment plant stability. This is because the standard deviation is theoretically a measure of dispersion, and practically it is the most familiar and simplest term among all other measures of stability.

All different measures of stability for 43 treatment plants were computed, and the linear relationship between them was examined using the Pearson correlation coefficient, a single number that summarizes the relationships between two variables. (30) The results of this analysis are shown in Table 11. A strong linear correlation exists between standard deviation and mean, maximum, and coefficient of stability ($r \geq 0.92$) indicating that standard deviation will also represent variations in mean and maximum values. This correlation is not significant for other indicators.

Comment

From the results reported in Table 11, it can be concluded that generally a high linear relationship ($r = 0.82$) exists between the annual means, standard deviations, and maximums of effluent BOD and SS concentrations. Thus, in most cases, when a plant is unstable in terms of biological treatment (BOD removal), it is expected to be unstable in terms of physical treatment (SS removal) too, and vice versa.

The Effect of Outliers on Stability Measures

Some selected time-series plots of effluent BOD and SS concentrations are shown in Appendix A. The computer printouts represent a two-dimensional plot of effluent concentration versus time. From an examination of these plots, it can be seen that in many plants there are one, two, or three days in one year of operation

TABLE 11. PEARSON CORRELATION COEFFICIENTS[a], r, OF ANNUAL DATA FOR 43 ACTIVATED SLUDGE TREATMENT PLANTS[b]

		BOD						SS					
		X	S_X	V_X	Max	COS	Antilog (LnS_X)	X	S_X	V_X	Max	COS	Antilog (LnS_X)
BOD	X	1.00	0.92	0.19	0.80	0.74	0.17	0.81	0.67	0.30	0.56	0.50	0.34
	S		1.00	0.53	0.94	0.94	0.43	0.82	0.82	0.54	0.74	0.72	0.60
	V			1.00	0.61	0.74	0.78	0.38	0.64	0.71	0.66	0.71	0.72
	Max				1.00	0.93	0.36	0.80	0.87	0.64	0.82	0.81	0.63
	COS					1.00	0.58	0.75	0.87	0.68	0.83	0.84	0.72
SS	X							1.00	0.88	0.34	0.81	0.75	0.34
	S								1.00	0.65	0.97	0.97	0.58
	V									1.00	0.71	0.74	0.80
	Max										1.00	0.98	0.57
	COS											1.00	0.61

[a] The Pearson correlation coefficient, r, for variables X and Y is defined as:[8]

$$r_{x,y} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{S_x} \right) \left(\frac{Y_i - \bar{Y}}{S_y} \right)$$

[b] All measurements are at 0.001 significance level

that effluent BOD or SS concentration is much higher than the rest of the year. This situation may strongly affect the results of both the annual mean and variance. Standard deviation from such data may be misleading when used as a stability measure, since it may be dominated by a few (one to three) samples (called outliers) that have extremely high values. The plant may be classified as unstable because of a few outliers, whereas, if the outliers are neglected, the plant might be classified as stable. These outliers may be caused by a mechanical failure of part of the system, stormy weather, and strong wind over the clarifiers, errors in sampling measurements, or any other unexpected cause.

Statistical Examination of the Effect of Outliers on Stability--To examine the effects of outliers on plant performance, statistical properties of 99 percentile values (99%ile) of the data were calculated and are given in Table 12. In each plant, one percent of the upper data values (3 days in one full year of operation) was removed from the data set, and the statistical properties of the rest of the data (99%iles) were calculated. Comparison of the results of Table 12 with Table 2 (statistical properties of 100%iles of the data) confirms the strong effect of outliers on the data set.

Removal of the upper one percent of the data resulted in a significant reduction in the maximum values. Annual mean effluent BOD and SS concentrations reduction were on the average of 0.5 and 1.0 g/m^3, respectively. Standard deviation of effluent BOD and SS concentrations was reduced on the average of 1.6 and 3.6 g/m^3, respectively. Reduction of the means and standard deviations of effluent SS concentration were remarkably high and generally higher than BOD. This finding can be taken as an indication that the relative magnitude of outliers exist in effluent SS data is greater than in BOD. Thus, process upsets will relate more often to high effluent SS values than high effluent BOD values.

As discussed later in this section, the outliers do not have any significant effect on the stability comparison of the plants (Tables 14 and 15). That is, the ranking of plants in terms of stability is roughly the same with and without outliers. Although the actual standard deviation values differ significantly when the upper one percent of the data (outliers) is removed, the standard deviation of 100%ile of the data is still an appropriate measure of stability.

Removing the upper one percent of the data set may not be convenient for operators or acceptable to regulatory agencies. Only if high value data points can be proven to be true outliers, that is, the result of accidents or measurement errors, should their removal be considered. These data observations may be true values of process performance and be characteristics of effluent concentration distributions. The nature of these outliers is determinable only if the source and reason for each outlier was noted and documented properly by the plant operator. Because the lack of appropriate reasons for rejection of these high value data points and because they generally do not affect the stability comparison of the plants, in this study these data points will be kept and treated as true values of effluent quality performance. These outliers should be rejected only if there is some reason for rejection and not solely because their value is high. Since these outliers are found in almost all plants, it is likely that they are characteristics of this type of data. If such values are true outliers, the causes or sources must be a characteristic of the type of test, equipment, and instrumentation used or the training of treatment plant operators.

TABLE 12. STATISTICS OF 99%ILE DATA OF EFFLUENT BOD AND SS CONCENTRATION

Plant Number	BOD Mean g/m^3	BOD S g/m^3	BOD Max g/m^3	SS Mean g/m^3	SS S g/m^3	SS Max g/m^3
1	9.40	6.30	30.0	10.03	8.78	62.0
2	8.30	4.08	21.0	6.20	2.38	14.0
3	9.73	6.02	39.0	14.47	4.52	26.0
4	17.78	10.15	57.0	25.06	18.96	113.0
5	15.78	8.28	46.0	20.06	11.40	86.0
6	16.12	8.39	41.0	20.79	11.92	98.0
7	20.53	9.38	72.0	14.92	12.98	110.0
8	15.11	9.56	56.0	18.99	18.66	123.0
9	23.79	19.65	94.0	42.58	43.95	220.0
10	29.79	10.37	65.0	57.90	16.75	114.0
11	17.53	14.10	97.0	8.97	10.95	78.0
12	16.21	6.24	42.0	14.06	8.74	54.0
13	15.59	12.51	62.0	14.17	14.84	75.0
14	21.48	15.85	93.0	31.40	33.67	211.0
15	25.33	19.76	105.0	20.24	28.62	160.0
16	4.79	2.30	14.0	5.22	2.57	13.0
17	1.94	1.39	7.0	2.41	2.46	13.0
18	2.27	0.78	4.0	4.45	2.06	11.0
19	4.73	1.91	10.0	8.09	3.64	19.0
20	7.63	3.62	18.0	6.70	2.93	23.0
21	4.30	2.07	14.0	6.15	2.88	18.0
22	4.23	1.96	11.0	5.60	2.57	17.0
23	2.92	1.59	9.0	3.52	1.82	10.0
24	23.24	25.02	112.0	17.35	18.90	95.0
25	12.50	8.17	47.0	10.83	6.78	54.0
26	8.86	4.00	25.0	9.32	5.38	28.0
27	12.32	6.06	32.0	8.73	5.23	26.0
28	13.51	6.08	29.0	8.64	6.16	38.0
29	4.10	3.38	16.0	9.83	5.49	29.0
30	5.88	5.51	24.0	9.68	6.10	30.0
31	10.57	7.87	48.0	10.62	3.91	24.0
32	23.88	9.70	46.0	23.77	8.65	52.0
33	17.05	13.90	60.0	20.68	13.36	78.0
34	8.64	4.87	24.0	14.40	6.67	38.0
35	29.86	21.07	155.0	33.95	34.23	177.0
36	6.60	4.54	19.6	7.12	3.50	22.0
37	20.43	10.99	55.0	18.73	14.13	58.0
38	15.15	15.04	79.0	59.82	67.97	412.0
39	84.84	56.42	305.0	76.96	66.31	400.0
40	14.20	5.00	29.0	8.63	4.87	27.0
41	8.98	3.77	22.0	9.53	6.22	34.0
42	20.01	12.06	90.0	26.37	30.00	233.0
43	19.73	19.67	170.0	38.58	57.52	376.0

Time-Series Analysis and Stability

An overview on the time-series plots of effluent BOD and SS concentration (Appendix A) illustrates that in some plants a type of annual trend or cycle exists. For example, Plants No. 10, 29, and 30 have a quadratic cycle, and Plants No. 4, 35 may have linear or logarithmic trends. Other plants show no annual trend or cycle at all.

The time dependency of the data must be recognized in calculating the overall mean and variance of the data. If a cycle or trend exists, the calculation of the dispersion of data should be measured from the curve of best fit through the data rather than from the mean, to reflect the effect of any cycle or trend of data into variance calculations. Variance is a measure of deviations from a horizontal line (mean) through the data and does not measure the effect of any trends or cycles in the data.

To examine the effect of the data time dependency, the dispersion of data in each plant was estimated by the mean square error (MSE) of the curve of best fit through the data. (The MSE is mean of the squares of the differences between each data point and the line of best fit.) Polynomials of degree less than or equal three, and other curves (exponential, hyperbolic) were fitted to the data, and the curve of best fit was determined. Standard deviation was estimated by the square root of the mean square error term of the best fitted curve ($S_x = \sqrt{MSE}$). Results from the estimation of standard deviation by mean square error of linear and parabolic curves are shown in Table 13. Generally, the parabolic curves result in the minimum mean square error. The cubic polynomial usually gives a lesser MSE than the square polynomial, but the improvement is not significant.

By comparing the results of standard deviation from Table 13 and Table 2, it can be concluded that generally variations will be reduced by recognizing the time effect of the data, but with few exceptions (Plants No. 10, 14, 15, 24, 29, 30, 35), the improvement is not significant. In other words, in activated sludge plants the effect of trends or cycles is negligible in comparison with the effect of other sources of variability. Since the data contain only one year of operation, there is no evidence that plants having a given cycle form will follow that form in succeeding years of operations too. These arguments lead to the conclusion that the mean square error is not a better indicator of stability than standard deviation.

Ranking Treatment Plants with Respect to Their Stability

In Tables 14 and 15, all plants have been ranked between 1 and 10 in ascending order with respect to their stability as measured by standard deviation of 100%iles and 99%iles of the data, by square root of the mean square error ($\sqrt{MSE}$), and by the coefficient of stability. In ranking, plants with the same or similar values of standard deviation (or other stability measure values) have been ranked in one group. Each step increase in ranking has been chosen based on a sharp change in the value of the standard deviation. It should be noted that this ranking is completely arbitrary, and its purpose was to compare the different stability measures.

TABLE 13. ESTIMATION OF STANDARD DEVIATION BY MEAN SQUARE ERROR, (S $=\sqrt{MSE}$), OF BEST FITTED LINE THROUGH EFFLUENT BOD AND SS CONCENTRATIONS

Plant Number	BOD		SS	
	S_x Linear	S_x Parabola	S_x Linear	S_x Parabola
1	6.45	6.12	12.76	12.78
2	4.31	4.26	2.66	2.65
3	6.44	6.33	4.47	3.87
4	10.63	10.40	19.99	19.94
5	8.55	8.28	15.69	15.70
6	8.64	8.34	16.09	16.10
7	11.37	11.16	21.25	21.12
8	15.63	15.32	26.18	25.63
9	22.17	21.93	48.42	47.36
10	11.62	9.03	18.42	15.36
11	16.34	16.16	17.42	17.43
12	6.55	6.39	11.19	10.77
13	14.09	13.58	16.63	16.22
14	21.18	18.46	45.94	45.29
15	19.58	18.46	30.72	28.51
16	2.48	2.38	2.54	2.53
17	1.09	0.84	2.55	2.12
18	0.77	0.77	1.94	1.93
19	1.71	1.72	3.94	3.93
20	3.41	3.35	4.33	4.20
21	2.37	2.37	3.58	3.49
22	2.01	2.01	3.04	3.00
23	1.63	1.36	2.03	1.91
24	22.49	21.87	23.87	23.31
25	8.91	8.84	12.71	12.63
26	4.38	4.24	6.35	6.34
27	6.32	6.28	5.32	5.12
28	6.58	6.57	6.83	6.70
29	3.61	2.51	5.65	4.44
30	5.29	4.39	6.38	5.38
31	8.56	8.51	4.73	4.70
32	9.32	8.80	9.08	9.08
33	15.60	15.60	16.42	16.02
34	5.20	5.13	7.50	7.21
35	24.48	24.31	33.32	32.97
36	4.62	4.44	3.84	3.84
37	12.43	12.45	11.49	11.25
38	18.45	18.13	83.15	82.64
39	57.72	56.81	73.98	74.08
40	4.75	4.76	4.85	4.21
41	3.78	3.78	6.48	6.44
42	13.51	12.92	36.78	36.49
43	26.16	26.06	69.23	69.29

TABLE 14. RANKING OF WASTEWATER TREATMENT PLANTS WITH RESPECT TO THEIR STABILITY, BASED ON THE STANDARD DEVIATION OF 100%ILE AND 99%ILE MEAN SQUARE ERROR, AND COEFFICIENT OF STABILITY OF EFFLUENT BOD CONCENTRATION

Stability ranking	Plant number	S 100%ile	Plant number	S 99%ile	Plant number	$\sqrt{MSE}$	Plant number	COS
1	18	0.78	18	0.78	18	0.77	18	0.27
	17	1.39	17	1.39	17	0.85	19	0.85
	23	1.71	23	1.59	23	1.36	22	0.97
	19	2.02	19	1.91	19	1.72	23	0.98
	22	2.03	22	1.96	22	2.01	17	1.00
	21	2.38	21	2.07	21	2.37	21	1.28
	16	2.51	16	2.30	16	2.38	16	1.29
2	20	3.76	29	3.38	29	2.51	41	1.63
	29	3.80	20	3.62	20	3.35	20	1.83
	41	3.84	41	3.77	41	3.78	40*	1.96
3	2	4.37	26	4.00	26	4.24	26	2.13
	26	4.39	2	4.08	2	4.26	2	2.25
	36	4.68	36	4.54	30	4.39	12*	3.00
	34	5.24	34	4.87	36	4.44	34	3.13
	40	5.31	40	5.00	40	4.76	28*	3.14
	30	5.68	30	5.51	34	5.13	27*	3.33
							36	3.26
							29*	3.37
4	27	6.47	3	6.02	1	6.12	32*	4.04
	28	6.58	27	6.06	27	6.28	10*	4.44
	3	6.67	28	6.08	3	6.33	3	4.45
	12	7.05	12	6.24	12	6.39	6*	5.21
	1	7.30	1	6.30	28	6.57	5*	5.29
							30*	5.39
							1	5.47
5	31	8.48	31	7.87	5	8.28	7*	6.23
	25	9.03	25	8.17	6	8.34	25	6.34
	5	9.27	5	8.28	31	8.51	31	6.64
	6	9.28	6	8.39	32	8.80		
	32	9.85			25	8.84		
					10*	9.03		

TABLE 14 (continued)

Stability ranking	Plant number	S 100%ile	Plant number	S 99%ile	Plant number	$\sqrt{MSE}$	Plant number	COS
6	4	11.40	7	9.38	4	10.40	4	7.09
	7	11.50	8*	9.56	7	11.16	37	7.76
	10	11.61	32*	9.70				
	37	12.77	4	10.15				
			10	10.37				
			37	10.99				
7	42	14.07	42	12.06	37	12.45	42	9.57
	13	14.18	13	12.51	42	12.92	13	12.32
					13	13.58		
8	8	16.10	33	13.90	8	15.32	33	14.55
	33	16.19	11	14.10	33	15.60	11	14.90
	11	16.59			11	16.16	8	15.86
9	38	19.84	38	15.04	38	18.13	15	19.74
	14	21.83	14	15.85	14	18.46	9	20.33
	9	22.44	9	19.65	15	18.46	14	20.77
	15	22.90	43*	19.67	24*	21.87	35*	23.28
			15	19.76	9	21.93	38	23.86
10	24	27.00	35	21.07	35	24.31	24	29.93
	35	27.14	24	25.02	43	26.06	43	34.52
	43	27.18	39	56.42	39	56.81	39	39.27
	39	57.96						

[*] Plants that have different ranking based on two different stability indicators

TABLE 15. RANKING OF WASTEWATER TREATMENT PLANTS WITH RESPECT TO THEIR STABILITY, BASED ON THE STANDARD DEVIATION OF 100%ILE AND 99%ILE, MEAN SQUARE ERROR, AND COEFFICIENT OF STABILITY OF EFFLUENT SS CONCENTRATION

Stability ranking	Plant number	S 100%ile	Plant number	S 99%ile	Plant number	$\sqrt{MSE}$	Plant number	COS
1	23	2.03	23	1.82	23	1.91	18	1.01
	18	2.13	18	2.06	18	1.93	23	1.14
	2	2.69	2	2.38	17	2.12	3*	1.46
	17	2.88	17	2.46	16	2.53	2	1.56
	16	2.93	16	2.57	2	2.65	16	1.60
	22	3.09	22	2.57	22	3.00	22	1.65
	21	3.67	21	2.88	21	3.49		1.88
	36	3.87	20*	2.93	36	3.84		
	19	3.93	36	3.50	3*	3.87		
			19	3.64		3.93		
2	20	4.47	31	3.91	20	4.20	36*	2.06
	3	4.60	3	4.52	40*	4.21	31	2.10
	31	4.79	40*	4.87	29*	4.44	21*	2.12
					31	4.70	20	2.85
3	40	5.28	27	5.23	27	5.12	40	3.16
	27	5.51	26	5.38	30	5.38	17*	3.24
	29	5.80	29	5.49	26	6.34	29	3.37
	26	6.37	30	6.10	41	6.44	27	3.42
	41	6.79	28	6.16	28	6.70	32*	3.56
	30	7.09	41	6.22	34	7.21	34	4.00
	28	7.20	34	6.67			26	4.19
	34	7.69	25*	6.78			41	4.70
							30	5.00
							28	5.77
							10*	5.95
4	32	9.26	32	8.65	32	9.08	12	9.08
	12	11.55	12	8.74	12	10.77		
	25	12.70	1	8.78	37*	11.25		
	1	12.82			25	12.63		
					1	12.78		

TABLE 15 (continued)

Stability ranking	Plant number	S 100%ile	Plant number	S 99%ile	Plant number	$\sqrt{\text{MSE}}$	Plant number	COS
5	37	14.61	11	10.95	10	15.36	37	11.17
	5	15.67	5	11.40	5	15.70	5	11.60
	6	16.11	6	11.92	33	16.02	6	11.84
	33	16.38	7*	12.98	6	16.10	33	12.33
	13	16.72	33	13.36	13	16.22	25*	13.55
	11	17.54	37	14.13	11	17.43	1*	15.01
	10	18.71	13	14.84				
6	7	21.23	10	16.75	4	19.94	4	18.25
	4	21.87			7	21.12	13*	18.97
7	24	25.09	8	18.66	24	23.31	7*	27.05
	8	26.53	24	18.90	8	25.63	11*	29.81
			4*	18.96			24	33.34
							8	33.55
8	15	34.03	15	28.62	15	28.51	35	40.54
	42	37.06	42	30.00	35	32.97	15	41.11
	35	37.92			42	36.49	42	48.34
							9*	53.24
9	14	47.75	14	33.67	14	45.29	14	65.63
	9	48.66	35*	34.23	9	47.36	39	69.23
			9	43.95				
10	43	70.87	43	57.52	43	69.29	43	118.74
	39	74.46	39	66.31	39	74.08	38	121.42
	38	88.86	38	67.97	38	82.64		

*plants that have different ranking based on two different stablity indicators.

Examination of the results presented in Tables 14 and 15 leads to the conclusion that rejecting outliers (i.e., the upper one percent data values) and/or fitting a curve through the data to remove the time-effect on data does not significantly change the stability ranking of the plants used in this study. All changes in ranking are shown with an asterisk in Tables 14 and 15. Therefore, all four indicators of stability shown in Tables 14 and 15 may be an acceptable measure of stability. But among all possible statistical indicators of stability "standard deviation" is the most appropriate because it is based on sound statistical theory. The fact that the standard deviation is simple to use, the familarity of users with the term, and because it does not have some of the weaknesses associated with other indicators discussed earlier in this section are additional reasons for its use.

STABILITY CUT-OFF POINT

Stability is usually defined in relative terms. For example, in comparing the stability of two treatment plants, the plant with less effluent quality variation is considered more stable. Stability may also be defined in absolute terms where a valid time deterministic model is available. For instance, if the monthly variance of effluent quality of a plant continuously increases each month (or any other time period), then, that plant is an unstable one in absolute terms. When variance changes randomly, the plant may or may not be absolutely unstable. A major factor would be the amount of the variance. Other alternatives are that the plant may be unstable for short periods of time or may simply be less stable than other plants with better performance characteristics.

The stability cut-off point may be defined as a standard deviation value below which plants are considered stable and above which plants are considered unstable. In Figures 11 and 12, the mean, standard deviation, and range of effluent BOD and SS data for all 43 treatment plants are shown. Examination of these descriptive statistics leads to the conclusion that plants having a standard deviation of less than 10 g/m^3 for both effluent BOD and SS may statistically be considered as "stable" plants. The 10 g/m^3 value has been selected as the stability cut-off point because a distinct difference exists between the statistical characteristics of the plants operating below and above this value. This cut-off point can be visualized clearly by examining Figures 11 and 12. Stable plants usually operate with mean effluent BOD values less than 17 g/m^3 and mean effluent SS values less than 15 g/m^3. Two-thirds of the time their effluent BOD and SS concentrations are less than 25 g/m^3 and 20 g/m^3, respectively. The daily concentration of effluent BOD and SS in these plants rarely exceeds 60 g/m^3. These plants have been ranked between 1 and 5 in Table 14 based on effluent BOD and ranked between 1 and 3 in Table 15 based on effluent SS concentration.

Plants with a standard deviation of greater than 10 g/m^3 for both effluent BOD and SS may be considered "unstable". Mean effluent BOD and SS values for these plants are generally greater than for the plants labeled stable, but some overlap occurs. Unstable plants have annual mean BOD and SS values as low as 16 g/m^3 and 10 g/m^3, respectively. The maximum daily BOD and SS concentrations of these plants are usually much greater than stable plants and exceed 90 g/m^3 and 130 g/m^3, respectively. These plants are ranked from 6 to 10 in Table 14 based on effluent BOD and ranked from 4 to 10 in Table 15 based on effluent SS concentration.

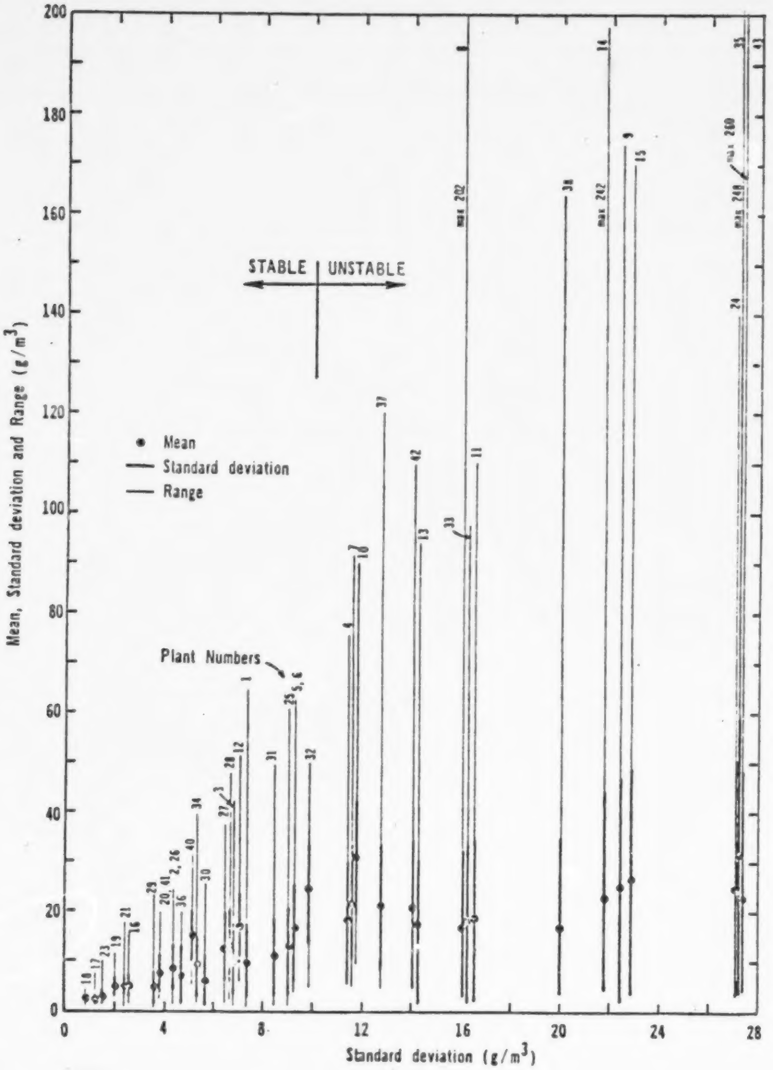


Figure 11. Mean, standard deviation, and range of effluent BOD concentration in different treatment plants.

EPA Form 2350-4 (1-80)
TYPING GUIDE SHEET

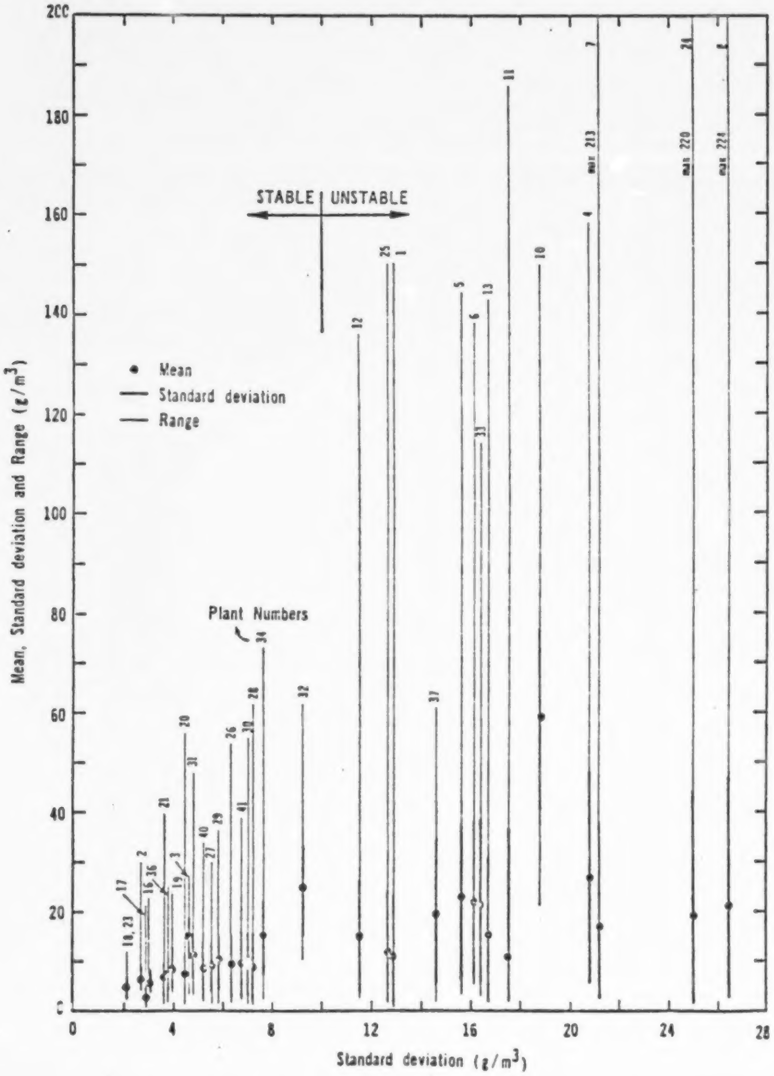


Figure 12. Mean, standard deviation, and range of effluent suspended solids concentration in different treatment plants.

Examination of the results of this analysis also leads to the conclusion that generally effluent SS shows more variation than effluent BOD and is expected to be a more significant factor in determining instability. Process upsets may happen more often due to high values of SS than to high values of BOD in the final effluent. To achieve the same level of stability for both BOD and SS, a plant should be designed to produce an average effluent SS less than effluent BOD.

THE EFFECT OF PLANT SIZE ON STABILITY

An assumption is usually made that larger treatment plants will produce better quality effluents than smaller plants. This assumption is based on the idea that larger plants are typically more sophisticated and have more highly trained operators. Some functional relationships in which the effect of size is dampened would be expected because plants of a certain size or larger would have equivalent sophistication. Hovey et al (23) in their study of optimal size of regional wastewater treatment plants concluded that there is a poor correlation (and possibly negative) between plant size and mean effluent concentration. They did not determine if an optimum size plant (i.e., one that produces the best quality effluent) exists, however.

Correlations between annual mean flow (as an indication of size of plant) with annual mean and standard deviation of effluent BOD and SS were determined to examine the significance of plant size on effluent quality and stability. Correlation coefficients of $r = 0.19$ and 0.17 were determined between annual mean flow with annual mean effluent BOD and SS, respectively; and coefficients of $r = 0.23$ and 0.16 were determined with standard deviations of effluent BOD and SS, respectively. These results also confirm the poor correlation between plant size and effluent quality; that is larger plants do not necessarily produce better quality effluent than smaller plants. Generally, there appears to be no relationship between plant size and stability.

THE EFFECT OF PROCESS TYPE ON STABILITY

Activated sludge processes are flexible systems and have been adopted to a number of different flow schematics. Ideally, the most appropriate process design will reduce the variability of effluent quality significantly. In conventional plug-flow activated sludge process, because both influent and recycled sludge enter the tank at the head end, effects of flow and concentration variations or the introduction of toxic substances is maximized and may result in upsets. However, any longitudinal mixing would tend to smooth out fluctuations in these variables. (15)

Step-aeration or step-feed processes are a modification of conventional activated sludge process to even out the organic load or the oxygen demand in a treatment process. (34) This approach gives an approximation of a complete-mix process. This modification may result in the reduction of effluent BOD and SS concentration and thus, production of a more stable process, since it will reduce some in-plant variations and also the effect of shock loads to the plant.

In complete-mix systems, the influent sewage and return sludge flow are introduced at several points in the aeration tank and mixed virtually instantaneously throughout the system. The organic load on the aeration tank and the oxygen demand are uniform throughout the tank. (28) In this case, any input fluctuations or toxic concentrations will be diluted instantaneously with the entire contents of the aeration unit. This kind of process, thus, would be expected to be more stable.

The contact-stabilization process is the best candidate for treating wastewaters containing very low soluble organic concentration and high non-soluble organic materials. (34) Part of the system is a physical or biosorption mechanism and is therefore less temperature sensitive and requires less aeration volume than the biooxidation process.

Statistical Examination of the Effect of Process Type on Stability

To examine the relationship between process type and stability, activated sludge treatment plants under study were grouped by process modification type and then were compared with each other. The data set included 18 conventional, 13 step-feed and step-aeration, 5 complete-mix, 4 contact-stabilization, one extended-aeration, one kraus, and one aerated lagoon. The average ($\bar{X}$) and standard deviation (S_x) of the annual means, standard deviations, and coefficients of variation of each type of process were computed. Results of the analysis are presented in Table 16 and can be used for comparison of the various modifications of the activated sludge process in terms of performance and stability.

Generally, step-feed or step-aeration modifications of activated sludge give the best year-round results for BOD removal, whereas, the conventional activated sludge gives the best year-round results for SS removal. In terms of BOD and SS removal, step-feed and conventional activated sludge processes operate closely and give similar year-round results; the complete-mix modification of activated sludge has somewhat lower performance characteristics; and the contact-stabilization process appears to have considerably lower performance characteristics than the other three systems. The other process types were not included since only data for one plant was available.

Examination of the results for each process leads to the conclusion that conventional and step-feed processes generally produce an effluent BOD and SS concentration below average and the complete-mix and contact-stabilization processes produce effluent BOD and SS concentrations greater than average when all plant data are pooled.

The Effect of Process Type on Stability Using a Non-Parametric Test--A non-parametric statistical test (the Mann-Whitney Test) (27) was used to evaluate the results presented in Table 16. In this test, the annual means and annual standard deviations of any two different activated sludge process types were compared. The objective was to determine if sufficient evidence had been presented to support a conclusion that a difference in the population distributions existed. The Mann-Whitney test is documented in Reference 27.

Using the Mann-Whitney test, it was also found that the step-feed and step-aeration processes generally produce the least effluent BOD and SS concentrations. Conventional activated sludge ranks second, complete-mix process ranks third, and contact-stabilization process ranks fourth.

TABLE 16. STATISTICS OF THE ANNUAL EFFLUENT BOD AND SS CONCENTRATION DATA FOR DIFFERENT PROCESS TYPES

Process type	No. of plants	Mean		Standard deviation		Coef. of variation[a]	
		$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}$	S_x
				BOD			
Conventional	18	12.80	6.85	9.54	7.99	0.69	0.25
Complete-mix	5	16.82	6.67	13.24	6.51	0.77	0.13
Step-Feed/Aeration	13	10.84	7.68	8.28	7.56	0.68	0.25
Contact Stabilization	4	38.38	32.08	28.17	20.92	0.76	0.14
Extended Aeration	1	14.41	—	5.31	—	0.37	—
Kraus	1	24.02	—	9.85	—	0.41	—
Aerated Lagoon	1	30.35	—	11.61	—	0.38	—
All Plants	43	15.76	13.43	11.28	10.49	0.70	0.23
				SS			
Conventional	18	14.92	10.53	16.02	18.61	0.86	0.38
Complete-Mix	5	19.88	14.19	19.65	16.92	1.00	0.51
Step-Feed/Aeration	13	16.23	16.65	16.83	23.89	0.83	0.34
Contact-Stabilization	4	40.88	26.75	37.66	26.18	0.90	0.14
Extended-Aeration	1	8.82	—	5.28	—	0.60	—
Kraus	1	24.12	—	9.26	—	0.38	—
Aerated Lagoon	1	58.79	—	18.71	—	0.32	—
All Plants	43	19.40	17.03	18.35	20.60	0.84	0.37

[a] Average values from all plants, not calculated from parameter averages

Discussion

Examination of the field data as described above contradicts generally held concepts of the relative stability of different process types. Contrary to the general belief, contact-stabilization and complete-mix processes do not perform in a more stable manner than conventional activated sludge processes. The reason for these results was not investigated here. However, possible causes include the effects of unmeasured variables such as environmental conditions, overloading, or operational and human factors.

THE EFFECT OF PROCESS TYPE ON RELIABILITY

Results shown in Table 16 may be used for comparison of the various modifications of activated sludge for the purpose of estimating process reliability. The expected mean and coefficient of variation for each process type may be drawn from Table 16. These values can be used in Figure 7 to estimate the process reliability. For example, the estimated reliability for various process types, assuming the upper limit of 30 g/m^3 for effluent BOD and SS concentrations are shown in Table 17. Generally, step-feed and conventional processes are more reliable followed by the complete-mix process and then by contact stabilization process.

STABILITY BASED DESIGN OF ACTIVATED SLUDGE PROCESSES

Analysis of data, as discussed earlier in this section, has resulted in the conclusion that plants having a standard deviation of less than 10 g/m^3 for both effluent BOD and SS concentrations may be statistically considered as stable plants. The question is, what mean value of effluent BOD and SS concentration may result in a stable process, while considering variations in effluent quality? Variations in effluent quality are described by the coefficient of variation, V_x. This coefficient may be estimated for each process type using the results of Table 16 or 17.

To produce a stable effluent quality (i.e., a process with a standard deviation of equal to or less than 10 g/m^3), the process should be designed for a mean value equal to or less than $\bar{X} = S_x/V_x$, where $S_x = 10$ g/m^3 and V_x may be obtained from Table 17. These values for different process types are shown in Table 18. To achieve stability, activated sludge processes should be designed for an effluent BOD concentration equal to or less than 13 to 15 g/m^3 and an effluent SS concentration equal to or less than 10 to 12 g/m^3.

TABLE 17. ESTIMATED RELIABILITY OF VARIOUS PROCESS TYPES, ASSUMING STANDARD LEVEL OF X_s = 30 g/m

Process Type	Average, $\bar{X}=m_X$ g/m	V_x	m_X/X_s	Reliability
BOD:				
Step Feed/Aeration	10.84	0.68	0.36	0.98
Conventional	12.80	0.69	0.43	0.94
Complete-Mix	16.82	0.77	0.56	0.88
Contact-Stabilization	38.38	0.76	1.28	0.50
SS:				
Step Feed/Aeration	16.23	0.83	0.54	0.88
Conventional	14.72	0.86	0.50	0.70
Complete-Mix	19.88	1.00	0.66	0.63
Contact-Stabilization	40.88	0.90	1.36	0.49

TABLE 18. RECOMMENDED DESIGN VALUES FOR EFFLUENT BOD AND SS CONCENTRATIONS IN ORDER TO PRODUCE A STABLE PROCESS

Process type	BOD $\leq$ (g/m)	SS $\leq$ (g/m)
Step-Feed/Aeration	14.70	12.05
Conventional	14.50	11.63
Complete-Mix	13.00	10.00
Contact-Stabilization	13.16	11.11

SECTION 6

FACTORS AFFECTING EFFLUENT VARIABILITY FROM ACTIVATED SLUDGE PROCESSES

Variations in the effluent quality from activated sludge processes are the result of a number of internal and external factors. In recent years, a considerable research effort has been directed toward examining process performance variations. (1,3,5,15,16,18,21,22,32,44,45) Both experimental and analytical research studies have been conducted; however, most of the experimental studies have been carried out under laboratory rather than field conditions. Thus the results reported do not necessarily reflect conditions encountered in full-scale operations.

OBJECTIVE

The objective of this part of the study was to determine the nature and the extent of the variability in effluent quality and to identify factors contributing to this variability through statistical analysis of operating data. Some potential benefits of the results of this study include: improved understanding of process performance variability which will allow for the design of more stable and reliable processes.

NATURE AND SOURCES OF EFFLUENT VARIABILITY

Several factors contribute to the perturbations observed in effluent quality. The following factors have been identified as being important.

1. Influent variables
2. Environmental conditions
3. Biological and operational parameters
4. Settling characteristics and clarifier design
5. Size of plant and type of process
6. Time effects
7. Human factors
8. Inherent process variability

While some of the above factors are controllable, others not, but all influence the quality of the effluent. The effect of plant size, process type, and time were discussed in the previous section. A brief background about other major factors are discussed here.

Influent Variables

The quality and quantity of the input loads into the treatment system change with time and usually cannot be controlled. Variations in the waste characteristics, including organic concentration and flow as well as the degradability of organic matter, and presence of toxic materials may influence effluent quality directly. (3)

Three types of time variable waste loading may be considered (32): 1) sudden or step changes, 2) periodical variations, and 3) random fluctuations. Sudden changes are caused by a batch of pollutant discharge into the system and a step input is caused by a sudden drop or increase in input flow and concentration such as start-up or shut-down of a process or heavy rainfalls. Diurnal, weekly, and seasonal oscillations of input load are typical examples of the periodical input variations. Efficiency of sewer collection system and the seasonality of some major industrial effluent flows are some other reasons for flow and organic load fluctuations. (41)

An increase in influent flow to the treatment plant will decrease the retention time and increase the pollutant load in the system. This causes flow surges in the system, turbulence, and a higher overflow rate in the final clarifier, which in turn results in a higher discharge of solids over the effluent weir. (3)

Changes in the influent organic loading will often affect the effluent concentration. However, a universal model describing relationships between influent and effluent organic concentrations with other variables has not been developed. In some studies based on laboratory research, it was found that with a constant organic loading, the effluent soluble BOD is directly proportional to the influent BOD. (16) On the other hand, in mass balance models based on completely mixed and steady-state conditions, the effluent organic concentration is independent of the influent organic concentration. (28,34)

In addition to the perturbations in raw waste loading, variability may also result from the biodegradability of the wastewater. Many wastes vary significantly in biodegradability, thereby affecting treatment process. Many industrial facilities have multi-product operations, and alterations in their product output may result in changes in their wastewater characteristics. For example, vegetable and fruit canneries often process a number of products during season with resulting changes in wastewater flow rate and characteristics. Many industrial wastewaters are significantly different from plant to plant as well as product to product.

Environmental Conditions

Environmental conditions, such as temperature and wind, may affect process performance.

Temperature--Organic removal rate is highly dependent upon temperature. The effect of temperature on the organic removal rate coefficient has been defined by the relationship: (28)

$$K_T = K_{20^\circ C}\,\theta^{(T-20)} \tag{23}$$

where K_T and K_{20} are the removal rate coefficients at T and 20°C, respectively. The value of the coefficient θ has been reported (28) to vary from 1.056 in the temperature range between 20 and 30°C to 1.135 in the range between 4 and 20°C. A temperature decrease of only 5°C can result in a decrease in the soluble BOD removal rate of about 30 to 80 percent.

In addition to the effect of temperature on soluble BOD removal, Adam (3) suggests that a decrease in temperature will result in poor settling, too. During the low temperature period, the viscosity of the water will increase, and the flocs tend to break easier; both conditions result in a poor settling of solids' flocs.

Wind--Another environmental effect on effluent quality is the influence of wind on secondary clarifier performance. Wind agitates the surface of large clarifiers and results in secondary currents forming that hinder the settling of light flocculent solids. (43) Such an effect is direct and results in increased suspended solids in the effluent. Often, a few high effluent concentration values are reported during the year which appear to be inconsistent with the data of the rest of the year. It is possible that wind effects may be the cause of these anomalous values.

Biological and Operational Parameters

In a well designed and operated process, the influent load fluctuations (BOD and SS) should not affect the effluent quality significantly. Because of the changes in influent organic concentration, the operator of a plant adjusts the biomass concentration in the system to maintain the mean cell residence time (MCRT) or food to microorganism ratio (F:M) within a range that results in a satisfactory effluent quality.

The biomass concentration adjustment will be made through operational procedures such as sludge return and sludge wasting. Return of activated sludge from the clarifier to the aeration tank provides sufficient concentration of biomass in the aeration tank. Wasting the excess of sludge is required to maintain a given F:M ratio or a constant MCRT. (28) In activated sludge process, a wide operating region, with respect to MCRT, exists in which the effluent quality variation is small. This range is between 3 to 10 day MCRT; below which the effluent organic concentration begins to rise due to deflocculation of particles and wash-out of predators in the system; and above which a large fraction of cell mass becomes non-respiring organic materials, which ends up with floc breakup and loss of solids over the weirs. (34)

Since plant operators adjust the operating parametes (e.g., sludge wastage and recycle rate), the biological and operational variables are not constant, but vary from day-to-day.

Settling Characteristics and Clarifier Design

One of the main sources and causes of variation in effluent quality is associated with the secondary clarifier and settling characteristics of sludge. The function of the clarifier is to separate the activated sludge solids from the mixed liquor. This is an essential part of the activated sludge process if the production of a high quality, stable effluent, low in BOD and SS is to be obtained. Examination of the levels and variations in effluent BOD and SS, as discussed in the previous section, leads to the conclusion that the sludge separation units are the limiting factors in production of a satisfactory high quality effluent.

A common factor in poor sludge settleability is bulking sludge. Bulking is associated with the presence of significant numbers of filamentous bacteria in the flocs. Such flocs have low density and compactability and therefore, do not settle well. Decreased settleability characteristics of the sludge results in a higher concentration of effluent suspended solids in the clarifier.

Sludge Volume Index (SVI) is used as an empirical indication of the settling characteristics of the sludge. Values of the SVI varies with the characteristics and concentration of the MLSS; thus, observed SVI values at a given plant cannot be compared with other plants. (28) In general, SVI values that are higher than the value normally observed can be taken as an indication that the sludge is not settling well. In this study, SVI and settleable solids were used as indicators of sludge settling characteristics. The extent of data available is not enough for examination of other factors such as type and size of the tank, surface, and solid loading rate used in the design of the solid separation facilities. These factors will be examined later in the continuation of the study.

Human Factors and Variability

System design criteria, loading factors, estimations and projections, and the capability of control overriding systems are among the many human factors that can significantly influence the capability of a process to produce a high quality effluent. (43) Even in well designed processes with sophisticated equipped control systems, human factors play an important role. Many sources of effluent quality perturbations may be controlled or reduced through careful operation of the system. Good operation and maintenance procedures are necessary to reduce or eliminate unexpected variations caused by equipment failures or other mechanical problems. Examination of the effects of human factors on effluent quality is not within the scope of this project, but should be examined in future studies.

Measurement Errors--Sampling procedures and analytical techniques are among the most significant sources of variation and errors. Measurement errors do not affect receiving water quality directly but do affect the ability to manage water quality. For example, if a process effluent BOD is consistently underestimated as the result of a procedural error in analysis, process performance could be judged satisfactory in terms of meeting secondary standards but unsatisfactory in terms of protecting receiving water quality because of conditions in the stream. Measurement errors that are not consistent, result in increasing the apparent process variability. A stable process could be judged unstable because of measurement errors.

Sampling procedures such as location, handling, collection of samples, method of analysis, and technical experience of the analyst can cause alterations in the analytical results. (43) Grab samples are highly variable and are not reliable indicators of any parameter. Twenty-four hour flow-weighted composite sampling is a better way for measuring any parameter.

Standard procedures have been adopted for sampling and laboratory analyses to reduce the errors due to analytical techniques and to produce more uniform and reliable results. Error or variability of some tests is still large, however. For example, according to Standard Methods (42), the coefficient of variation of BOD and SS measurements are about 6.5 and 5 percent, respectively.

Reproducibility between replicate samples is largely a function of sample size and experience of the analyst. Usually all sample results should fall within a 5 percent range, however.

Inherent Variability

The "inherent variability" is the variability in effluent quality associated with a properly designed and well-operated process. Inherent variability is the result of the basic nature of the treatment process and cannot be altered significantly by external forces. (18) Wastewater characteristics, geographical and climatological conditions, and the nature of biota in the treatment system are among the sources of this type of variability. Where industrial operations contribute to the wastewater flow, degradability of the organic materials may be a factor. Other factors such as short circuiting in aeration or sedimentation tanks and species interactions can contribute to the inherent variability of a process.

In this study, factors such as oxygen supply, pH, and wind were included as sources of inherent variation because available data did not allow examination of their effects. Possible nutrient deficiencies must also be included as a source of inherent variation. Oxygen, pH, and nutrients are factors that should be considered in any process design. Wind effects can be mitigated by baffling or covering sedimentation tanks.

STATISTICAL ANALYSIS OF EFFLUENT VARIATIONS

A suitable approach in developing a baseline variability for effluent quality is to analyze field data from operating plants. Several statistical concepts may be used to develop such a baseline for analyzing and understanding fluctuations and the relationships between different variables. As the first step of data analysis, the distributional characteristics of each variable must be determined. The distributional characteristics provide the necessary information to analyze variability as well as information required for selection of subsequent statistical techniques. (30)

Descriptive Statistics

Distributional characteristics of selected variables and parameters were determined for all of the plants studied. Statistics developed from the field data are shown in Table 19. In addition, data for effluent BOD and SS are given in Table 2, Section 3.

At the time of this phase of the analysis, data from only 21 plants were completed and compiled for analysis. However, as discussed later in this section, development of the descriptive statistics or other statistical relationships for all plants would not lead to additional insight into the relationships. Thus, because of economic reasons and time requirements, data from other plants were not investigated. Results derived from Table 19 are considered in the following discussion.

Flow--In general, the coefficient of variation for flow was relatively low for the 21 plants, ranging between 0.03 and 0.39. Flow fluctuation was characterized by a right skewed distribution in most cases.

Influent BOD and SS Concentration--The coefficients of variation for influent BOD and SS concentrations were between 0.21 and 0.77 and between 0.18 and 1.08, respectively. These two variables were also characterized by a right skewed distribution.

MCRT--The coefficient of variation for MCRT was high in many plants; as a result, it must be concluded that the values of the MCRT used for operation are too high. A typical curve (28,34) of effluent BOD as a function of MCRT is shown in Figure 13. As shown, MCRT values greater than 10 days have very little effect on effluent BOD changes. Therefore, high MCRT values that were often reported have little or no relationship to effluent quality variability. For example, in Plant No. 1 (high MCRT), the variability in the MCRT cannot be used the explain the effluent BOD variations. Among the plants studied, only a limited number were operated in the range where the MCRT has an effect on effluent quality.

MLSS--Average MLSS values varying from 1130 to 5550 g/m^3 were reported for the 21 treatment plants. The coefficient of variation for MLSS ranged between 0.12 and 0.36. The range of MLSS variations was low compared to the other biological and operational variables. Based on this finding, it would appear that an attempt is made to keep the MLSS concentration at a constant level.

Operational Parameters--The coefficient of variation for return sludge flow (Q_R) and concentration (X_R) was between 0.03 and 0.46 and between 0.12 and 0.40, respectively. The corresponding coefficient for waste sluge flow (Q_w) ranged between 0.15 and 1.49.

TABLE 19. STATISTICS OF SOME SELECTED VARIABLES FOR DIFFERENT TREATMENT PLANTS

Plant Number	Flow, mgd Mean	S_X	V_X	Skew	Max	Min	Valid Obs.	Wastewater Temperature (WTEMP), °F Mean	S_X	V_X	Skew	Max	Min	Valid Obs.
1	14.2	5.5	0.39	1.01	34.8	4.1	366	59.9	8.84	0.15	0.05	75	45	363
2	50.1	1.7	0.03	-7.25	56.0	27.0	366	75.4	2.94	0.04	0.03	82	70	352
3	16.3	1.8	0.11	-0.11	20.1	11.1	366	76.1	3.19	0.04	-0.29	82	66	358
4	23.7	3.0	0.13	0.09	33.0	16.6	364	58.7	6.27	0.11	0.10	69	45	364
5	69.7	12.4	0.18	0.74	105.5	44.4	366	64.5	8.51	0.13	0.18	79	50	366
6	66.2	11.9	0.18	0.76	104.0	41.6	366	64.5	8.51	0.13	0.18	79	50	366
7	69.3	15.1	0.22	0.94	118.3	41.1	366	64.5	8.51	0.13	0.18	79	50	
8	15.2	1.8	0.12	0.93	24.9	9.4	366							
16	6.8	1.5	0.22	0.28	9.7	3.9	212	71.7	4.00	0.06	0.53	80	63	209
17	2.4	0.2	0.09	-0.80	3.1	1.7	366	76.4	3.46	0.05	-0.02	83	70	364
18	29.1	2.6	0.09	-1.10	3.5	19.0	366	77.9	3.83	0.05	-0.14	85	70	365
19	7.1	0.6	0.08	0.05	8.7	5.4	366	76.9	4.30	0.06	-0.51	84	65	364
31	3.8	1.1	0.29	1.63	10.7	2.1	366	61.5	6.84	0.11	0.04	72	51	336
33	2.8	0.8	0.27	3.28	7.8	1.9	366	62.6	6.62	0.11	0.02	73	50	366
34	17.5	1.4	0.07	1.68	26.0	14.7	365	63.2	7.69	0.12	-0.97	75	30	259
38	17.3	2.8	0.16	1.91	32.0	10.1	364	63.6	6.63	0.10	0.02	76	51	357
39	33.3	3.8	0.12	-1.35	40.0	14.9	365							
40	0.6	0.1	0.26	1.11	1.1	0.4	365	56.3	8.01	0.14	-0.09	69	44	251
41	26.9	8.8	0.33	2.27	75.6	13.1	299	57.8	8.62	0.15	-0.46	69	37	266
42	81.6	13.8	0.17	-0.07	123.1	41.2	366	65.8	7.00	0.11	0.08	79	50	365
43	54.0	13.6	0.25	0.15	93.4	22.3	366	64.6	7.25	0.11	0.09	78	47	365

TABLE 19 (continued)

Plant Number	Influent BOD (BOD_i), g/m^3 Mean	S_X	V_X	Skew	Max	Min	Valid Obs.	Influent SS (SS_i), g/m^3 Mean	S_X	V_X	Skew	Max	Min	Valid Obs.
1	54	29.2	0.55	1.95	256	9	334	85	54.5	0.64	2.80	490	12	365
2	256	54.2	0.21	1.82	580	150	360	284	68.9	0.24	4.25	956	196	360
3	345	101.5	0.29	2.02	1200	160	364	475	147.6	0.31	3.06	1730	220	365
4	176	42.4	0.24	0.20	316	79	365	136	24.8	0.18	0.92	278	74	366
5	189	103.6	0.55	2.54	840	52	365	293	214.9	0.73	8.87	3420	64	366
6	189	103.6	0.55	2.54	840	52	365	293	214.9	0.73	8.87	3420	64	366
7	189	103.6	0.55	2.54	840	52	365	293	214.9	0.73	8.87	3420	64	366
8	281	215.3	0.77	1.48	1054	53	260	552	595.0	1.08	2.10	3482	47	260
16	206	68.6	0.33	0.13	474	35	211	240	77.3	0.32	1.33	686	22	211
17	301	99.0	0.33	0.48	534	112	73	400	144.8	0.36	2.00	1324	200	358
18	355	97.3	0.27	1.16	775	152	81	392	105.6	0.30	4.74	1623	178	363
19	268	65.8	0.25	1.75	578	137	83	315	137.3	0.44	3.17	1288	108	339
31	138	56.8	0.41	1.04	330	47	150	132	44.3	0.34	1.76	452	51	364
33	193	60.7	0.31	0.60	410	58	152	322	123.8	0.37	0.88	752	94	157
34	203	82.4	0.41	1.13	670	60	334	273	97.8	0.36	2.00	941	100	365
38	149	52.5	0.35	2.28	440	72	364	180	99.5	0.55	4.12	1172	64	364
39	394	96.7	0.25	0.43	773	151	365	330	174.0	0.53	3.19	1841	80	365
40	162	42.2	0.29	1.03	345	78	251	137	56.8	0.41	1.34	380	48	251
41	76	25.3	0.33	0.33	154	15	304	142	71.2	0.50	2.84	658	27	298
42	326	96.0	0.30	0.38	680	120	361	312	86.9	0.28	0.48	651	88	363
43	326	96.0	0.30	0.38	680	120	361	312	86.9	0.28	0.48	651	88	363

TABLE 19 (continued)

Plant Number	Mixed Liquor Suspended Solids (MLSS), g/m^3 Mean	S_X	V_X	Skew	Max	Min	Valid Obs.	Food to Microorganism Ratio (F:M) Mean	S_X	V_X	Skew	Max	Min	Valid Obs.
1	5551	1295	0.23	0.00	10063	460	365	0.12	0.11	0.89	3.11	0.85	0.02	315
2	2749	406	0.15	-0.25	3710	1750	363	0.28	0.07	0.24	0.48	0.54	0.11	333
3	1651	272	0.17	0.64	3000	650	363	0.37	0.12	0.34	0.27	0.71	0.13	361
4	1342	232	0.17	0.90	2320	852	366	0.33	0.10	0.31	0.24	0.62	0.11	366
5	2933	636	0.22	0.57	5765	1100	366							
6	3245	767	0.24	0.71	6145	1025	366							
7	3209	533	0.17	1.63	6240	2140	365							
8	4219	1240	0.29	0.57	9926	851	260	0.15	0.08	0.50	2.35	0.73	0.03	261
16	1554	560	0.36	0.92	3413	643	204							
17	1671	379	0.23	0.31	2700	822	365							
18	2012	266	0.13	-0.17	2723	1241	366							
19	1215	149	0.12	0.61	1758	864	365							
31	1912	551	0.29	0.71	3476	889	366							
33	2605	355	0.14	-0.30	3810	902	366							
34	1130	187	0.17	0.64	1930	596	362							
38	3165	1093	0.35	0.18	8020	930	365	0.53	0.32	0.60	5.63	4.25	0.15	365
39	5180	1518	0.29	0.01	10020	1830	363	0.38	0.16	0.41	1.64	1.11	0.12	362
40	2896	729	0.25	0.46	4570	1515	362	0.17	0.05	0.31	0.81	0.36	0.07	251
41	1718	559	0.33	0.49	3900	530	298	0.36	0.17	0.48	1.96	1.29	0.08	283
42	3769	569	0.15	1.36	63	2700	366	0.33	0.11	0.31	0.15	0.69	0.08	347
43	3810	589	0.16	1.17	6100	2700	366	0.28	0.11	0.39	0.50	0.80	0.06	348

TYPING GUIDE SHEET

TABLE 19 (continued)

Plant Number	Mean Cell Residence Time (MCRT), Day							Hydraulic Residence Time (HRT), Hour						
	Mean	S_X	V_X	Skew	Max	Min	Valid Obs.	Mean	S_X	V_X	Skew	Max	Min	Valid Obs.
1	30.20	87.84	2.91	11.36	1233.1	4.9	247	4.93	1.81	0.37	1.08	13.6	1.8	366
2	4.78	0.82	0.17	1.46	10.0	2.7	258							
3	4.04	0.67	0.17	0.27	7.2	2.5	361	4.05	0.46	0.11	0.59	5.9	3.2	366
4	4.95	0.89	0.18	0.60	9.9	1.3	366	5.31	0.70	0.13	0.59	7.5	3.7	364
5								3.85	0.51	0.13	0.27	5.7	2.7	366
6														
7								2.99	0.43	0.14	-0.25	4.1	1.9	366
8														
16	22.47	35.66	1.59	4.38	293.1	4.8	212							
17	16.58	32.52	1.96	7.19	388.3	0.0	364							
18	35.56	37.81	1.06	3.42	271.0	6.7	366							
19	8.23	5.78	0.70	5.48	66.8	3.6	365							
31														
33	44.35	64.52	1.46	3.26	458.3	1.5	157							
34	3.11	1.72	0.55	3.33	17.0	0.7	362							
38	2.14	1.45	0.68	1.80	11.0	0.2	361							
39	2.62	1.15	0.44	3.48	13.3	0.8	362	4.29	0.77	0.18	3.39	9.3	3.1	290
40	9.05	3.17	0.35	0.12	20.2	3.0	233							
41	3.18	2.70	0.85	4.96	26.2	0.7	291							
42														
43														

TABLE 19 (continued)

Plant Number	Dissolved Oxygen (DO), g/m^3 Mean	s_X	v_X	Skew	Max	Min	Valid Obs.	Wastage Flow (Q_w), mgd Mean	s_X	v_X	Skew	Max	Min	Valid Obs.
1	2.29	0.68	0.30	2.02	6.60	1.10	366	0.05	0.03	0.63	1.08	0.19	0	248
2	0.56	0.24	0.43	1.02	1.40	0.20	255							
3	3.50	0.76	0.22	-0.13	6.40	1.20	359	0.39	0.07	0.18	0.49	0.60	0.18	363
4								1.98	0.34	0.18	-0.51	2.57	0.35	356
5	4.12	1.37	0.33	0.62	8.70	0.90	366							
6	3.63	1.46	0.40	0.80	9.20	1.20	366							
7	4.11	1.20	0.29	0.84	8.50	2.10	366							
8	2.75	1.24	0.45	0.70	7.20	0.40	259							
16	3.12	1.12	0.36	0.27	6.35	0.35	200							
17	1.25	0.68	0.54	1.09	4.15	0.30	221							
18														
19	3.26	1.00	0.31	-0.29	5.60	0.10	360							
31														
33	1.93	0.92	0.48	1.11	5.50	0.70	366	0.05	0.07	1.49	1.51	0.34	0	366
34														
38								1.04	1.32	0.26	2.11	5.91	0.03	364
39								0.89	0.15	0.17	-0.54	1.27	0.19	365
40								0.01	0	0.20	0.79	0.01	0	335
41								0.56	0.24	0.44	0.17	1.28	0	298
42								1.56	0.24	0.15	0.12	2.44	0.42	365
43								1.02	0.21	0.20	0.56	1.79	0.26	365

TABLE 19 (continued)

Plant Number	Return Sludge Concentration (X_R), g/m^3 Mean	S_X	V_X	Skew	Max	Min	Valid Obs.	Return Sludge Flow (Q_R), mgd Mean	S_X	V_X	Skew	Max	Min	Valid Obs.
1	21441	6592	0.31	2.17	72000	2000	365	6.73	2.11	0.31	0.69	18.59	1.78	365
2	8557	1113	0.13	0.49	11380	6220	259	23.45	4.32	0.18	0.35	33.30	14.30	366
3	6495	802	0.12	1.00	12230	4310	363							
4	4065	665	0.16	0.44	6486	2616	366	9.11	0.24	0.03	0.89	10.04	8.28	364
5	8634	2845	0.33	2.15	28100	3500	366	39.70	11.60	0.29	-0.26	92.80	6.50	366
6	8485	2587	0.31	1.96	26500	3900	365	25.12	3.86	0.15	-1.37	30.00	7.70	366
7	11974	2582	0.22	0.81	19900	5100	365	27.33	1.65	0.06	-0.30	31.50	21.90	366
8								6.68	0.86	0.13	-0.03	9.06	4.62	366
16	5481	2026	0.37	0.30	11172	1228	212	3.03	0.56	0.19	1.59	5.34	1.74	204
17	5262	1309	0.25	0.16	8558	2255	365	1.17	0.26	0.22	1.34	2.04	0.72	366
18	4802	727	0.15	0.13	6695	2927	366	4.16	0.32	0.08	-0.15	6.19	3.10	366
19	3738	511	0.14	0.49	5796	2513	366	2.93	0.39	0.13	1.54	4.39	2.21	366
31														
33	6547	1611	0.25	1.75	15852	2032	366	1.57	0.30	0.19	0.74	2.50	0.40	366
34														
38	7440	2302	0.31	0.53	16900	2100	364	11.03	1.41	0.13	0.41	14.80	7.40	365
39	10931	3662	0.40	-0.01	21800	2600	364							
40	9904	2174	0.22	0.40	1754	5690	362	0.19	0.09	0.46	1.16	0.45	0.06	365
41	5841	2311	0.40	0.77	15120	840	298	9.85	2.96	0.30	1.94	23.85	4.52	291
42	13845	2259	0.16	0.64	21400	9200	366	28.57	3.59	0.13	0.52	40.30	16.80	366
43	13763	1663	0.12	0.25	19300	9200	366	19.39	3.37	0.17	0.11	28.90	10.80	366

TABLE 19 (continued)

Plant Number	Sludge Volume Index (SVI) Mean	S_X	V_X	Skew	Max	Min	Valid Obs.	Settleable Solids (SETSOL), ml/l Mean	S_X	V_X	Skew	Max	Min	Valid Obs.
1	51	12.7	0.25	2.03	143	26	362							
2	121	47.0	0.39	1.68	315	52	365	13.08	3.71	0.28	3.14	33	8	359
3	112	22.6	0.20	0.98	246	43	361	35.03	17.18	0.49	2.16	150	8	339
4	279	94.2	0.34	0.21	528	89	366							
5	83	20.4	0.25	0.44	142	40	366							
6	85	23.4	0.28	1.23	184	26	366							
7	82	20.2	0.25	1.14	184	36	365							
8	153	113.9	0.74	3.80	917	54	359							
16	177	90.0	0.51	0.78	431	56	205	236.90	135.60	0.57	1.15	750	50	205
17	182	88.2	0.48	2.44	685	40	366	335.50	140.40	0.42	0.71	830	50	366
18	180	57.4	0.32	1.25	496	90	364	316.40	102.40	0.32	1.15	705	140	366
19	156	80.4	0.51	2.81	694	23	365	173.90	81.90	0.47	2.63	720	20	366
31	200	165.5	0.83	1.98	838	71	364							
33	189	82.8	0.44	0.79	398	60	366	121.70	53.80	0.44	1.08	324	10	157
34														
38	76	26.9	0.36	2.84	270	23	365	22.30	9.40	0.42	0.33	62	6	365
39	89	33.5	0.38	1.10	200	16	362							
40	113	24.2	0.21	1.44	217	75	362	336.50	140.50	0.42	1.52	885	170	362
41	58	16.1	0.28	0.37	112	25	299	9.57	3.31	0.35	0.31	19	3	301
42	128	38.3	0.30	0.72	263	60	310							
43	104	31.6	0.30	1.15	222	55	310							

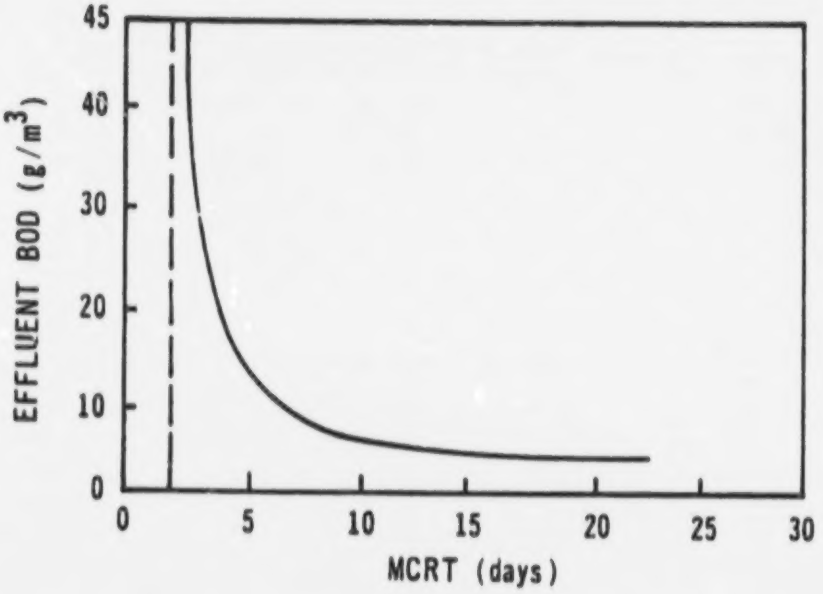


Figure 13. Effluent BOD as a function of Mean Cell Residence Time[27].

SVI--The coefficient of variation for SVI varied from 0.20 to 0.83. Maximum values of SVI over 300 were obtained in many plants.

Wastewater Temperature--Variation of wastewater temperature can be divided into two groups. Plants that are located in geographical areas with moderate changes in air temperature had a coefficient of variation varying from 0.04 to 0.06, whereas plants in colder areas had values between 0.11 to 0.15. The average wastewater temperatures for the first group of plants were in high 70's and for the second group were in high 50's and low 60's.

Discussion

High values and wide ranges of the coefficient of variation for different variables are the indication of high degree of dispersion in all plants and in each variable. Generally, the relative magnitude of the coefficient of variation between several variables for one plant can be used as an indicator of the possible sources of significant variation in each plant.

For example, in Plant No. 38, the coefficient of variation for effluent BOD and SS (Table 2) are 1.20 and 1.37, respectively. Therefore, a high degree of dispersion exists in the effluent quality performance of this plant. The coefficient of variation for input loads (Flow, BOD_i, and SS_i) are low, from this information it can be concluded that input loads may not be a major cause of effluent dispersion in this plant. The coefficients of variation for operational and biological parameters (MCRT, MLSS, F:M, Q_R, X_R) are all low to moderate; thus, these factors have little variations. Wastewater temperature variation is relatively high compared to other plants. Sludge volume index changes are moderate in comparison to other plants. Thus, it may be concluded that among the variables considered temperature is the only major cause of fluctuation in this plant and that the remaining of the variability must be classed as inherent.

In Plant No. 1, effluent BOD and SS variations are also high. Input loads have a relatively high coefficient in variation, and thus, may be a cause of the perturbations in the effluent quality. In such a plant, flow equalization may be of use in reducing effluent variability and improving the stability of the plant. Wastewater temperature changes for this plant are considerably higher than the other plants, and that could be a cause for a variation in this plant, also.

CORRELATION ANALYSIS

Correlation analysis has been employed to investigate the extent to which variations in effluent quality are linked to fluctuations in the other variables. The Bivariate Pearson correlation coefficient, symbolized by r, serves a dual purpose. First, it is used as an indicator of the goodness of fit of a linear regression, and second it is used as a measure of association indicating the strength of the linear relationship between two variables. (30) Because the correlation coefficient is normalized, the range of values is between 1 and -1. When a linear relationship between the variables does not exist $r = 0$, while a perfect correlation would result in $r = 1$. Thus, calculated r values near 1 are used to support assumptions of linear relationships. Positive correlations ($r > 0$) mean that the two variables increase or decrease together, and negative correlations ($r < 0$) denote an inverse relationship.

The square of the correlation coefficient, r^2, called coefficient of determination, is a measure of the proportion of variance in one variable that can be explained by the other. Because r has values between 1 and -1, r^2 is limited to the region from zero to one. Coefficient of determination is a more easily interpreted measure of association when concern is with the strength rather than sign of a relationship.(30)

Linear Correlation

The results of correlation analysis for the 21 treatment plants are presented in Tables 20 and 21. Coefficients of correlation between effluent concentration and other variables (e.g., influent or operational) were linearly insignificant in most cases and vary both in strength and sign. In some cases, effluent quality was related directly to one variable, while in another case, the same two variables were related inversely. Therefore, a single variable cannot be used to characterize the variability of effluent BOD and SS in all of the plants.

Influent Variables--A consistent relationship between flow and effluent concentration did not exist. In addition, influent BOD concentration did not have a significant linear relationship with effluent BOD. In most cases, less than 25 percent of effluent BOD variation (e.g., $r^2 = 0.25$) can be explained by changes in the influent BOD. Even less correlation was found between influent and effluent SS concentrations. The maximum r^2 value for the latter variables was 0.09.

Temperature--Wastewater temperature has an inverse correlation with effluent quality. This relationship was strong in some plants and was very weak in others. In some plants, up to 38 percent of effluent variation in BOD and 27 percent of effluent variation in SS is accounted for by the wastewater temperature.

Biological and Operational Parameters--In general, the relationship between biological and operational variables and effluent quality was very dispersed. A linear relationship between MCRT and effluent quality was not found, although within the range of 3 to 10 day MCRT, such a relationship would be expected. The lack of correlation between MCRT and effluent quality is an indication that MCRT was not being used (or was incorrectly used) as a control parameter. Maximum r^2 values between MCRT and effluent BOD and SS were 0.21 and 0.33, respectively. For most of the plants, these values were less than 0.10 for both variables.

Food-to-microorganisms ratio (F:M) was also not linearly related to effluent quality. Maximum r^2 values between F:M and effluent BOD and SS concentrations were 0.13 and 0.07, respectively.

Less than 14 percent of effluent BOD and SS variations were explained by changes in the MLSS. This variable had a positive correlation with effluent concentration in some plants and a negative correlation in other plants.

TABLE 20. CORRELATION ANALYSIS

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT BOD AND SELECTED VARIABLES

Plant Number	Variable											
	SS_e	Flow	BOD_i	WTEMP	MCRT	F:M	MLSS	X_R	Q_R	Q_w	DO	SVI	Time
1	0.54	0.15	0.03	-0.57	-0.01	0.01	0.03	-0.01	0.19	-0.09	0.24	0.03	-0.47
2	0.51	-0.06	0.14	-0.10	0.29	-0.14	0.37	0.03	0.53	-	0.08	0.52	-0.18
3	0.49	-0.07	0.20	-0.37	-0.03	0.25	-0.21	-0.10	-	-0.16	-0.21	-0.02	-0.27
4	0.82	-0.18	0.34	0.39	-0.46	0.36	-0.14	-0.21	-0.20	-0.21	-	0.34	0.36
5	0.62	-0.24	0.55	-0.09	-	-	0.09	-0.14	0.25	-	-0.31	0.31	0.39
6	0.60	-0.28	0.49	-0.04	-	-	-0.08	0.01	-0.10	-	-0.28	0.33	0.37
7	0.67	-0.07	0.40	-0.15	-	-	0.08	-0.07	0.14	-	-0.10	0.21	0.16
8	0.34	0.26	0.26	-	-	0.35	-0.19	-	0.12	-	0.17	0.16	0.25
9	0.26	0.42	0.48	-0.55	-0.23	-	0.32	0.43	-0.17	-	-0.27	-0.33	-0.19
10	0.65	-0.60	-0.27	-0.62	0.38	-	0.51	0.53	-0.25	-	0.21	-0.28	-0.63
11	0.53	0.33	0.16	-0.07	-0.12	-	-0.07	0.01	0.25	-	-	-0.29	-0.06
12	0.43	-0.11	-0.03	-0.34	-0.01	-	0.16	-0.05	0.29	-	-0.37	-0.19	-0.54
13	0.36	-0.04	0.05	-0.02	-	-	0.07	-	-	0.05	-	-0.07	-0.05
14	0.28	-0.08	0.41	0.27	-0.08	-	0.11	0.02	0.02	0.09	-0.13	-0.22	0.28
15	-0.03	0.02	-0.06	0.12	-0.08	-	-0.09	-	-	-	-	-	0.13
16	0.93	0.19	0.13	-0.40	-0.14	-0.01	0.38	0.28	0.10	-0.10	-	0.24	-0.37
17	0.41	0.01	0.34	-	-0.19	0.23	0.02	0.07	-	-0.02		0.02	-0.10
18	0.58	-0.09	0.36	-0.15	0.01	0.01	0.24	0.25	0.14	-0.05	-	0.11	-0.45
19	0.35	-0.09	0.33	-0.18	-0.14	0.31	-0.13	-0.14	-0.04	0.20	-	0.10	-0.19
20	0.84	0.06	0.25	-0.36	-	0.09	0.36	0.26	0.27	0.10	-	0.10	-0.29
21	0.95	0.17	0.04	-0.26	-	0.03	0.21	0.09	0.34	0.04	-	0.05	-0.27

TABLE 21. CORRELATION ANALYSIS

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT SS AND SELECTED VARIABLES

	Variable											
Plant Number	Flow	SS_i	WTEMP	MCRT	F:M	MLSS	X_R	Q_R	Q_w	DO	SVI	Time
1	0.14	0.18	-0.19	-0.11	-0.01	0.15	0.12	0.14	-0.08	0.08	-0.12	-0.11
2	-0.04	-0.03	-0.17	0.18	-0.20	0.32	0.03	0.39	-	0.09	0.42	-0.16
3	-0.12	-0.03	-0.51	-0.15	0.03	-0.30	-0.12	-	-0.10	-0.08	-0.33	-0.25
4	-0.20	0.25	0.34	-0.58	0.28	-0.12	-0.18	-0.14	-0.12	-	0.38	0.41
5	0.12	0.06	-0.07	-	-	-0.02	-0.09	0.05	-	-0.17	0.21	-0.01
6	0.10	0.06	-0.07	-	-	-0.19	0.06	0.07	-	-0.12	0.18	-0.07
7	0.14	0.15	-0.23	-	-	0.02	-0.10	0.21	-	-0.10	0.18	-0.03
8	0.22	0.11	-	-	0.17	-0.07	-	0.06	-	0.06	0.20	0.17
16	-0.31	0.23	-0.12	0.01	-	-0.24	0.10	-0.34	-	0.03	-0.57	0.50
17	-0.53	-0.08	-0.37	0.14	-	0.32	0.37	-0.13	-	0.38	-0.21	-0.46
18	0.42	0.04	-0.52	-0.23	-	0.03	0.23	0.12	-	-	-0.43	-0.41
19	0.01	-0.04	0.03	0.09	-	-0.08	-0.15	0.14	-	0.04	-0.12	-0.03
31	0.12	0.19	-0.20	-	-	0.05	-	-	-0.11	-	0.15	-0.16
33	0.19	0.18	-0.15	-0.36	-	0.10	0.21	0.04	-0.05	-0.21	-0.06	0.03
34	-0.13	0.09	-0.16	0.07	-	0.11	-	-	-	-	-	-0.23
38	0.22	0.11	-0.36	-0.15	-0.03	0.39	0.29	0.15	-0.12	-	0.18	-0.36
39	0.07	0.06	-	-0.41	0.17	-0.16	-0.19	-	0.06	-	0.16	0.12
40	-0.27	0.30	-0.22	0.05	-0.08	-0.03	0.08	0.19	-0.24	-	-0.08	-0.40
41	0.24	0.05	-0.34	-0.08	0.27	-0.14	-0.10	0.23	0.13	-	-0.26	-0.30
42	0.11	0.13	-0.22	-	0.04	0.18	0.05	0.37	0.11	-	0.15	-0.13
43	0.21	-0.04	-0.23	-	-0.02	0.16	0.06	0.39	0.13	-	0.05	-0.22

Among the operational variables, correlation between return sludge flow (Q_R), return sludge concentration (X_R), and waste sludge flow (Q_W) with effluent quality were determined. These variables were usually not significant factors in process alterations. The effect of these variables together is reflected in MCRT measurement, however.

The relationship between effluent quality and dissolved oxygen concentration in the aeration tank was also examined for some plants. Dissolved oxygen was not an important factor in explaining the effluent quality fluctuations.

Sludge Volume Index--Sludge volume index, SVI, had a relatively low correlation with effluent BOD and SS. Up to 30 percent of effluent deviations was explained by SVI.

Effect of Time--In several plants, a linear relationship between the time of the year and effluent quality was found. However, the correlations between effluent quality and time of the year compare directly to the correlations between effluent quality and wastewater temperature. This finding leads to the conclusion that, in most cases, the time relationship originates from temperature changes. When higher r values exist between the time of the year and the effluent concentration as compared to those between temperature and effluent concentration, it is possible that the difference may have resulted from operational or technological changes in the plant throughout the year. Lower correlations between effluent concentration and time (after subtracting the effect of temperature) may be interpreted to mean that the plant was operated in the same manner throughout the year.

Non-Linear Correlation

Non-linear correlations, including square, parabolic, and exponential relationships between effluent BOD or SS and other variables, were also examined. Some of the results are reported in Tables 22 and 23. Comparison of the results of the non-linear correlation with the results of linear correlation (Tables 20 and 21) leads to the conclusion that improvement, if any, is not that significant when using non-linear regression analysis. As in the case of linear correlation analysis, a single variable cannot be used to explain the variability of effluent BOD or SS in a non-linear manner.

The Effect of Lag Periods

Usually there is some lag between the time that influent waste enters the plant and the time effluent waste exits the plant. Retention times within plants would be expected to vary between 2 and 24 hours, with typical values in the range of 8 to 12 hours. Thus, the lack of significant correlations between effluent concentration and the selected variables might be because of the existence of such a lag.

Data presented here are from a 24-hour flow-weighted composite sampling. This procedure relaxes the effect of the lag between influent and effluent variables to a considerable amount. Although it is believed that composite sampling dampens

TABLE 22. NON-LINEAR CORRELATION ANALYSIS

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT BOD AND SELECTED VARIABLES

Plant Number	Variable[a]						
	Flow	BOD$_i$	WTEMP	MCRT	F:M	MLSS	SVI
1	0.12	0.06	-0.56	0.01	-0.01	0.02	0.04
	0.12	0.06	-0.56	0.01	0	0.02	0.04
	0.17	0.01	-0.58	-0.08	0.03	0.04	0.01
2	-0.07	0.11	-0.11	0.26	-0.17	0.35	0.50
	-0.07	0.11	-0.11	0.29	-0.15	0.35	0.50
	-0.04	0.16	-0.10	0.30	-0.11	0.38	0.51
3	-0.08	0.16	-0.37	-0.04	0.23	-0.20	-0.03
	-0.08	0.16	-0.37	-0.04	0.24	-0.20	-0.03
	-0.07	0.21	-0.37	-0.01	0.25	-0.21	-0.01
4	-0.19	0.33	0.39	-0.37	0.36	-0.12	0.32
	-0.19	0.33	0.39	-0.38	0.36	-0.12	0.32
	-0.17	0.34	0.39	-0.53	0.34	-0.17	0.34
5	-0.23	0.42	-0.10	-	-	0.08	0.31
	-0.23	0.42	-0.10	-	-	0.08	0.31
	-0.25	0.57	-0.09	-	-	0.08	0.30
16	0.42	0.46	-0.56	-0.18	-	0.25	-0.30
	0.42	0.46	-0.56	-0.18	-	0.25	-0.30
	0.41	0.44	-0.55	-0.32	-	0.37	-0.27
33	-0.05	0.45	0.27	-0.02	-	0.10	-0.22
	-0.06	0.45	0.27	-0.01	-	0.10	-0.22
	-0.09	0.34	0.26	-0.16	-	0.11	-0.20
38	0.20	0.09	-0.40	-0.13	-0.01	0.36	0.21
	0.20	0.09	-0.40	-0.13	-0.01	0.36	0.21
	0.18	0.18	-0.40	0.08	0	0.36	0.25
41	-0.08	0.29	-0.17	-0.07	0.24	-0.12	0.11
	-0.08	0.29	-0.17	-0.07	0.28	-0.12	0.11
	-0.09	0.34	-0.18	-0.20	0.34	-0.13	0.10

[a] For each plant the first line is square, the second line is parabolic and the third line is exponential relationship

TABLE 23. NON-LINEAR CORRELATION ANALYSIS

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT SS AND SELECTED VARIABLES

Plant Numbe	Variable[a]						
	Flow	SS_i	WTEMP	MCRT	F:M	MLSS	SVI
1	0.13	0.13	-0.19	0.07	0	0.16	-0.11
	0.13	0.13	-0.19	0.07	-0.01	0.16	-0.11
	0.13	0.19	-0.18	-0.07	0	0.12	-0.14
2	-0.05	-0.07	-0.18	0.16	-0.19	0.32	0.44
	-0.05	-0.07	-0.18	0.17	-0.20	0.32	0.44
	-0.03	0.01	-0.17	0.19	-0.20	0.32	0.38
3	-0.12	0.01	-0.51	-0.15	0.01	-0.28	-0.33
	-0.12	0.01	-0.51	-0.15	0.02	-0.28	-0.33
	-0.12	-0.07	-0.51	-0.14	0.05	-0.31	-0.33
4	-0.22	0.25	0.34	-0.48	0.27	-0.09	0.37
	-0.22	0.25	0.34	-0.49	0.28	-0.09	0.37
	-0.19	0.25	0.33	-0.65	0.28	-0.14	0.37
5	0.12	0	-0.08	-	-	-0.01	0.20
	0.12	0	-0.08	-	-	-0.01	0.20
	0.11	0.12	-0.06	-	-	-0.02	0.21
16	-0.30	0.22	-0.13	-0.09	-	-0.26	-0.46
	-0.30	0.22	-0.13	-0.09	-	-0.26	-0.46
	-0.32	0.20	-0.11	0.14	-	-0.20	-0.67
33	0.22	0.17	-0.15	-0.26	-	0.11	-0.04
	0.22	0.17	-0.15	-0.26	-	0.11	-0.04
	0.15	0.16	-0.16	-0.24	-	0.09	-0.09
38	0.23	0.06	-0.36	-0.13	-0.02	0.38	0.14
	0.23	0.06	-0.36	-0.14	-0.02	0.38	0.14
	0.21	0.13	-0.36	-0.08	-0.03	0.37	0.19
41	0.23	0	-0.33	-0.02	0.22	-0.11	-0.22
	0.23	0	-0.33	-0.02	0.24	-0.11	-0.22
	0.25	0.08	-0.35	-0.20	0.27	-0.16	-0.28

---

[a] For each plant the first line is square, the second line is parabolic and the third line is exponential relationship

the lag effects, correlations between effluent quality and the other variables were calculated to examine the effect of 24, 12, 6, and 3 hour lag periods. The results of this study are summarized in Tables 24 and 25. To examine the effect of a 24 hour lag, the daily values of the effluent BOD or SS were correlated with the values for the previous day. For 12 hour lag periods, each day effluent BOD or SS values were correlated with the average of the value of the same day and the previous day of the other variables. For 6 hour lag periods, each day's effluent BOD or SS was correlated with the sum of 75 percent of the same day value and 25 percent of the previous day value of each variable.

It is clear from a comparison of the results presented in Tables 24 and 25 with the results given in Tables 20 and 21 that the effect of the lag periods between influent and effluent variables is insignificant. In most cases, there is a stronger correlation between the effluent quality and other variables for the same day measurements than when considering the lag between these variables.

The Effect of Outliers

As discussed in the previous section, in many plants there exists one, two, or three days that effluent BOD or SS concentration is much higher than the rest of the year. These data points could be true values of effluent characteristics or could be outliers. The effect of these high value data points in correlation analysis was also examined. In each plant, one percent of the upper data values were removed before proceeding with the correlation analysis. The results are reported in Tables 26 and 27. By comparing these results with the results of the linear correlation (Tables 21 and 22), it can be seen than a higher correlation between some variables in some plants and a lower correlation between other variables in other plants exist. The differences in correlation of 100%ile and 99%ile data values are generally not significant enough to be considered further in this study.

REGRESSION ANALYSIS: EXAMINATION OF THE EFFECT OF SELECTED VARIABLES ON EFFLUENT QUALITY VARIATIONS

Regression analysis techniques have been used as a descriptive tool to evaluate the contribution of a specific variable or set of variables in the effluent BOD and SS variations. Selected groups of variables were studied, both individually and in combination, to determine the individual linear effect as well as combined effect of each group of variables on effluent fluctuations. Selected groups of variables relevant to the variability of daily effluent quality from the activated sludge treatment plants studied are:

1. Input loading: flow, influent BOD (BOD_i), and influent SS (SS_i)

2. Environmental: wastewater temperature (WTEMP).

3. Biological and operational parameters: Mean cell residence time (MCRT), Mixed-liquor suspended solids (MLSS), Food to microorganisms ratio (F:M), recycled sludge flow (Q_R) and concentration (X_R), and wastage sludge flow (Q_W).

TABLE 24. CORRELATION ANALYSIS WITH LAGS

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT BOD AND SELECTED VARIABLES

Plant Number	Variable[a]						
	Flow	BOD_i	WTEMP	MCRT	F:M	MLSS	SVI
1	0.15	-0.06	-0.60	0.02	0	0.03	0.05
	0.16	0	-0.59	0	0.01	0.03	0.04
	0.16	0.03	-0.58	-0.01	0.01	0.03	0.03
	0.15	0.05	-0.58	0.01	0.01	0.03	
2	-0.05	0.15	-0.12	0.36	-0.11	0.34	0.50
	-0.06	0.17	-0.11	0.39	-0.15	0.37	0.53
	-0.06	0.16	-0.11	0.35	-0.15	0.37	0.53
	-0.06	0.15	-0.11	0.33	-0.15	0.37	
3	-0.26	-0.11	-0.36	0.11	-0.12	-0.16	0.01
	-0.18	0.05	-0.37	0.04	0.07	-0.22	-0.01
	-0.13	0.14	-0.37	0.01	0.18	-0.24	-0.01
	-0.10	0.17	-0.37	-0.01	0.22	-0.22	
4	-0.25	0.20	0.36	-0.14	0.20	-0.07	0.36
	-0.23	0.33	0.38	-0.34	0.34	-0.11	0.36
	-0.21	0.32	0.39	-0.42	0.37	-0.13	0.36
	-0.20	0.35	0.39	-0.44	0.37	-0.14	
5	-0.32	0.45	-0.07	-	-	0.09	0.31
	-0.29	0.56	-0.08	-	-	0.10	0.34
	-0.27	0.57	-0.09	-	-	0.10	0.33
	-0.25	0.57	-0.09	-	-	0.09	
16	0.40	0.47	-0.54	-0.29	-	0.33	-0.38
	0.42	0.53	-0.55	-0.23	-	0.33	-0.36
	0.42	0.52	-0.55	-0.23	-	0.32	-0.35
	0.42	0.51	-0.55	-0.23	-	0.32	
33	-0.11	-	0.27	-	-	0.11	-0.18
	-0.10	-	0.27	-	-	0.14	-0.21
	-0.09	-	0.27	-	-	0.13	-0.22
	-0.08	-	0.27	-	-	0.12	

TABLE 24 (continued)

Plant Number	Variable[a]						
	Flow	BOD_i	WTEMP	MCRT	F:M	MLSS	SVI
38	0.04	0.07	-0.40	-0.01	-0.05	0.44	0.24
	0.12	0.14	-0.40	-0.08	-0.04	0.43	0.29
	0.16	0.14	-0.41	-0.11	-0.02	0.41	0.28
	0.18	0.14	-0.41	-0.13	-0.01	0.39	
41	-0.27	0.29	-0.17	-0.02	0.15	-0.19	0.09
	-0.21	0.36	-0.15	-0.09	0.29	-0.18	0.10
	-0.15	0.36	-0.16	-0.12	0.32	-0.17	0.11
	-0.12	0.35	-0.16	-0.13	0.32	-0.16	

[a] For each plant the first line 24 hr lag, the second line 12 hr, third line is 6 hr and fourth line is 3 hr lag time

TABLE 25. CORRELATION ANALYSIS WITH LAGS

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT SS AND SELECTED VARIABLES

Plant Number	Variable[a]						
	Flow	**SS_i**	**WTEMP**	**MCRT**	**F:M**	**MLSS**	**SVI**
1	0.12	-0.04	-0.21	-0.02	0.09	0.01	0
	0.14	0.09	-0.20	-0.07	0.08	0.08	-0.08
	0.14	0.16	-0.19	-0.09	0.04	-0.11	-0.11
	0.14	0.17	-0.18	-0.10	0.03	0.12	
2	-0.05	-0.10	-0.17	0.28	-0.25	0.30	0.40
	-0.05	-0.07	-0.18	0.25	-0.28	0.32	0.43
	-0.05	-0.07	-0.19	0.22	-0.25	0.32	0.43
	-0.04	-0.05	-0.18	0.20	-0.23	0.32	
3	-0.29	-0.15	-0.50	0	-0.17	-0.20	-0.28
	-0.22	-0.12	-0.51	-0.08	-0.08	-0.29	-0.33
	-0.17	-0.07	-0.51	-0.11	-0.02	-0.31	-0.34
	-0.14	-0.05	-0.51	-0.13	0	-0.30	
4	-0.26	0.07	0.31	-0.19	0.15	-0.03	0.37
	-0.25	0.22	0.33	-0.44	0.26	-0.08	0.39
	-0.23	0.26	0.34	-0.53	0.28	-0.10	0.39
	-0.22	0.26	0.34	-0.56	0.28	-0.11	
5	0.02	0.02	-0.06	-	-	-0.04	0.23
	0.07	0.05	-0.07	-	-	-0.03	0.24
	0.09	0.05	-0.07	-	-	-0.02	0.23
	0.11	0.06	-0.07	-	-	-0.02	
16	-0.32	0.12	-0.11	0.06	-	-0.24	-0.58
	-0.32	0.23	-0.10	0.04	-	-0.24	-0.59
	-0.32	0.25	-0.11	0.02	-	-0.24	-0.58
	-0.31	0.24	-0.11	0.02	-	-0.24	
33	0.14	-	-0.14	-	-	0.18	-0.10
	0.17	-	-0.15	-	-	0.18	-0.09
	0.18	-	-0.15	-	-	0.15	-0.08
	0.19	-	-0.15	-	-	0.13	

TABLE 25 (continued)

Plant Number	Variable[a]						
	Flow	SS_i	WTEMP	MCRT	F:M	MLSS	SVI
38	0.10	0.02	-0.46	0.01	-0.06	0.49	0.26
	0.17	0.09	-0.37	-0.08	-0.05	0.46	0.26
	0.20	0.11	-0.37	-0.12	-0.04	0.43	0.22
	0.21	0.11	-0.36	-0.13	-0.03	0.41	
41	-0.01	0	-0.30	0.04	0.15	-0.21	-0.22
	0.14	0.03	-0.32	-0.02	0.27	-0.18	-0.25
	0.20	0.04	-0.33	-0.06	0.29	-0.17	-0.26
	0.22	0.04	-0.33	-0.07	0.29	-0.15	

[a] For each plant the first line is 24 hr, the second line is 12 hr, the third line is 6 hr, and the fourth line is 3 hr lag time

TABLE 26. CORRELATION ANALYSIS OF 99%ILE OF DATA

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT BOD AND 99%ILE DATA OF SELECTED VARIABLES

Plant Number	Variable									
	SS_e	FLow	BOD_i	WTEMP	MCRT	F:M	MLSS	SVI	Q_R	X_R
1	0.40	0.22	-0.01	-0.64	0	0	0.04	0.02	0.27	-0.04
2	0.51	-0.06	0.15	-0.15	0.30	-0.16	0.38	0.55	0.57	0.02
3	0.47	-0.06	0.17	-0.35	-0.02	0.25	-0.18	-0.01	-	-0.10
4	0.77	-0.17	0.33	0.39	-0.40	0.34	-0.12	0.36	-0.21	-0.20
5	0.56	-0.27	0.55	-0.10	-	-	0.08	0.31	0.26	-0.11
9	0.32	0.42	0.52	-0.56	-0.23	-	0.29	-0.38	-0.21	0.45
13	0.18	-0.05	0.37	0.27	-0.11	-	0.04	-0.24	0.18	0.01
16	0.90	0.20	0.17	-0.43	-0.13	-0.01	0.40	0.20	0.07	0.32
19	0.37	-0.08	0.33	-0.16	-0.14	0.31	-0.13	0.12	-0.04	-0.13

TABLE 27. CORRELATION ANALYSIS OF 99%ILE OF DATA

PEARSON CORRELATION COEFFICIENT, r, BETWEEN EFFLUENT SS AND 99%ILE DATA OF SELECTED VARIABLES

Plant Number	Variable									
	BOD_e	Flow	BOD_i	WTEMP	MCRT	F:M	MLSS	SVI	Q_R	X_R
1	0.49	0.29	0	-0.32	-0.11	0	0.18	-0.20	0.30	0.12
2	0.52	-0.04	-0.06	-0.18	0.18	-0.22	0.34	0.43	0.41	0.04
3	0.48	-0.14	-0.02	-0.52	-0.13	0.01	-0.29	-0.34	-	-0.10
4	0.80	-0.20	0.25	0.37	-0.54	0.25	-0.16	0.41	-0.14	-0.22
5	0.55	0.02	0.13	-0.12	-	-	0.01	0.23	0.08	-0.04
9	0.27	-0.30	0.30	-0.10	0.03	-	-0.28	-0.59	-0.37	0.10
13	0.15	0.25	0.04	-0.21	-0.40	-	-0.01	-0.05	0.18	0.17
16	0.91	0.18	0.11	-0.40	0.12	-0.05	0.46	0.16	0.13	0.35
19	0.31	0.26	-0.13	-0.40	-0.08	0.27	-0.17	-0.28	0.23	-0.11

4. Settling characteristcs: sludge voume index (SVI) and settleable solids (SETSOL).

5. Time

6. All variables combined together.

Not all of the results of regression analysis are shown here. Extensive space would be required to present all of the regression analysis results, coefficients of regression lines and the analysis of variances associated with each study. However, summary results of the coefficient of determination, for each regression study for all plants are given in Tables 28 and 29.

The multiple coefficient of determination, R^2, is used to describe the strength of the dependence or the amount of variation in the effluent concentration that can be explained by linear depndence upon the independent variables operating jointly. (30) Thus, R^2 is analogous to r^2 for bivariate analysis. Note that in Tables 28 and 29 the R^2 values are reported, which directly can be used to explain the variability relationship rather than just correlation between the two variables as is explained by R.

In Tables 28 and 29, the most significant variables, those which explain at least 5 percent of effluent perturbations have been shown in parentheses. Other variables have less than 5 percent effect on effluent variability. In those plants that have one or more missing variables in their data reports, regression analysis cannot be carried out because the variable group is incomplete; thus, results are shown as a blank in these tables. The meaning of the coefficient of determination in Tables 28 and 29 is best explained by an example.

> For instance, in Plant No. 1, only 3 percent of effluent BOD changes can be explained by input variables (Flow, BOD_i). Neither flow nor influent BOD has notable effect on effluent BOD. Wastewater temperature accounts for 32 percent of the variation. Biological and operational parameters can be used to explain only 9 percent of effluent BOD fluctuations, where return flow Q_R, is the most important operational variable in this plant. Other biological or operational parameters do not contribute to the dispersion of effluent quality significantly. In this plant, SVI measurement does not account for the effluent BOD variations at all. Time of the year can be used to explain 22 percent of changes in effluent quality, which most probably is due to temperature deviations. A combination of all groups of variables can be used to describe 48 percent of effluent perturbations, and among all variables, wastewater temperature is the most important factor. The remaining variation (i.e., 52 percent) in this plant is not explained by the available set of data. The unexplained variation may be caused by other variables not included in the regression list, such as inherent variability, and human factors. Same explanation will hold for all other plants.

TABLE 28. REGRESSION ANALYSIS

COEFFICIENT OF DETERMINATION, R^2, FOR EFFLUENT BOD AND SELECTED GROUPS OF VARIABLES

Plant Number	Input[b]	WTEMP	Bio. & Oper[c]	SVI	Time	All Variables[d]
1	0.03	0.32	0.09 (Q_R)	0	0.22	0.48 (WTEMP)
2	0.03	0.01	0.29 (Q_R)	0.27	0.03	0.42 (SVI, MLSS)
3	0.05	0.14	-	0	0.07	0.21 (WTEMP)
4	0.16 (BOD_i)	0.15	0.25 (MCRT)	0.12	0.13	0.43 (MCRT,WTEMP,BOD_i)
5	0.30 (BOD_i)	0.01	-	0.10	0.15	-
6	0.25 (BOD_i)	0	-	0.11	0.13	-
7	0.17 (BOD_i)	0.02	-	0.04	0.02	-
8	0.11 (BOD_i)	-	-	0.03	0.06	-
9	0.34 (BOD_i,Flow)	0.30	0.21 (X_R)	0.11	0.04	0.40 (WTEMP)
10	0.37 (Flow)	0.38	-	0.08	0.40	0.61 (WTEMP,Time)
11	0.12 (Flow)	0.01	-	0.08	0	0.27 (Flow,Time)
12	0.01	0.12	-	0.04	0.29	0.40 (Time)
13	0	0	-	0	0	-
14	0.19 (BOD_i)	0.07	0.02	0.05	0.08	0.32 (BOD_i,SVI)
15	0	0.01	-	-	0.02	-
16	0.07	0.16	0.16 (MLSS)	0.06	0.14	0.32 (WTEMP)
17	0.12 (BOD_i)	-	-	0	0.01	-
18	0.14 (BOD_i)	0.02	-	0.01	0.20	0.34 (BOD_i,Time)
19	0.10 (BOD_i)	0.03	0.05	0.01	0.04	0.13 (BOD_i)
20	0.07 (BOD_i)	0.13	-	0.01	0.08	-
21	0.03	0.07	-	0	0.07	-

[a] Values in parentheses are the most important variable (greater than 5% significance) in explaining the total variation.

[b] Input loads: Flow, BOD_i. [c] Biological and operational variables: MCRT, MLSS, X_R, Q_R.

[d] All variables: Flow, BOD_i, WTEMP, MCRT, MLSS, SVI, Time.

TABLE 29. REGRESSION ANALYSIS

COEFFICIENT OF DETERMINATION, R^2, FOR EFFLUENT SS AND SELECTED GROUPS OF VARIABLES

Plant Number	Input[b]	WTEMP	Bio. & Oper.[c]	SVI	Time	All Variables[d]
1	0.05	0.04	0.11 (Q_R)	0.01	0.01	0.21 (Flow,SS_i)
2	0	0.03	0.17 (Q_R)	0.18	0.03	0.22 (SVI,MLSS)
3	0.01	0.26	-	0.11	0.06	0.41 (WTEMP,SVI)
4	0.10 (SS_i)	0.12	0.36 (MCRT)	0.14	0.17	0.52 (MCRT,SVI)
5	0.02	0.01	-	0.04	0	-
6	0.01	0	-	0.03	0	-
7	0.05	0.05	-	0.03	0	-
8	0.05 (Flow)	-	-	0.04	0.03	-
9	0.13 (Flow)	0.01	0.16 (Q_R)	0.32	0.25	0.52 (SVI,Time)
10	0.29 (Flow)	0.14	-	0.04	0.21	0.33 (Flow)
11	0.18 (Flow)	0.27	-	0.13	0.17	0.44 (WTEMP,Flow)
12	0	0	-	0.01	0	0.05
13	0.07 (SS_i)	0.04	-	0.02	0.03	-
14	0.11	0.02	0.20 (MCRT,X_R)	0	0	0.26 (MCRT,Flow)
15	0.03	0.03	-	-	0.05	-
16	0.06 (Flow)	0.13	0.17 (MLSS)	0.03	0.13	0.31 (MLSS,Time,MCRT)
17	0.01	-	-	0.02	0.01	-
18	0.12	0.05	-	0	0.16	0.36 (Time,Flow)
19	0.07 (Flow)	0.12	0.11 (Q_R, MLSS)	0.07	0.10	0.22 (WTEMP,Flow)
20	0.03	0.05	-	0.02	0.02	-
21	0.05	0.05	-	0	0.05	-

[a] Values in parentheses are the most important variable (greater than 5% significance) in explaining the total variation.

[b] Input loads: Flow, SS_i. [c] Biological and operational variables: MCRT, MLSS, X_R, Q_R.

[d] All variables: Flow, SS_i, WTEMP, MCRT, MLSS, SVI, Time

Discussion

The contribution of each group of variables on effluent quality variations can be derived from Tables 28 and 29 as discussed below.

Input Variables--Considering all plants studied, input variables accounted for a maximum of 37 percent of the effluent BOD fluctuations, and in most cases, their contribution was less than 17 percent. In a number of plants, influent BOD and flow were not the causes of the alterations in effluent BOD at all. Generally, the effect of influent BOD was more notable than flow, and with three exceptions, flow usually was not the responsible factor for effluent BOD variations.

The coefficient of determination between effluent SS and influent variables was generally low. The R^2 value for Plant No. 17 was 0.29, but all other R^2 values were below 0.18. Flow was found to be a more important factor than influent SS in describing effluent SS deviations. This result may be due to flow related effects on clarifier performance. In most cases, input variables were more responsible for effluent BOD variations than for effluent SS.

Wastewater Temperature--It was found that up to 38 percent of the effluent BOD variation and up to 27 percent of the effluent SS variation was caused by fluctuations in wastewater temperature. In many plants, temperature changes did not have any effect on effluent quality. The contribution of wastewater temperature in effluent variability was not dependent on the geographical area of the plant.

Biological and Operational Variables--Biological and operational variables (MCRT, MLSS, X_R, Q_R, Q_W) generally were not found to be significant factors in explaining the effluent quality variations. The total contribution of these variables to effluent BOD variations was less than 30 percent and in some cases, was close to zero. Their effect on fluctuations in effluent SS was generally greater than effluent BOD and was in the range of 11 to 36 percent.

Sludge Settleability--Sludge volume index (SVI) was generally reported as a measure of sludge settleability. Changes in SVI would be expected to be reflected in effluent quality variations, and this was true to some extent. Twenty-seven percent of the effluent BOD fluctuation was explained by SVI variation in Plant No. 2, but the r^2 values were less than 0.12 in all other plants. In a number of plants, there was no correlation at all between these two variables. Similar statements can be made about the correlations between SVI and effluent SS.

Combination of Variables--When the effects of all of the input, environmental, and operational variables were summed, between 13 and 61 percent of the effluent BOD and between 5 and 52 percent of the effluent SS fluctuations (depending on the plant) were explained. The remaining variations were unexplained (i.e., inherent or human factors as defined here). These results are summarized in Tables 28 and 29.

From these results, it is clear that the variables that affect plant performance vary from plant to plant. Any single variable or group of them may account for the fluctuations in effluent quality in different plants. For example, in Plant No. 1, wastewater temperature was the most important variable, while in Plant No. 2, SVI and MLSS contributed the greatest amount to fluctuations in effluent BOD. The contribution of the remaining variables was less than 5 percent.

Examination of the results presented in Tables 28 and 29 leads to the conclusion that there were only a few variables that contributed significantly to the changes in effluent quality. The number of times that each variable significantly contributed to the variations of effluent quality are shown in Table 30. Among all variables, wastewater temperature was the most notable variable in describing effluent variations.

The overall results of this phase of the study are insufficient to generalize about the variability of process performance. Inclusion of additional data, however, is unlikely to alter this conclusion, for a large dispersion exists among the plants that have been studied. Therefore, a specific analysis of data must be carried out at each plant to identify the main causes of variation.

Sources of Effluent Variation Unexplained by the Available Data--Many causes of performance variation exist that cannot be explained on the basis of analysis of the available data. Included among these variables are true inherent variability such as presence of toxic materials and nature of biota in the system and human factors including system design, operation and maintenance procedures, loading factors, systems control capability, and sampling techniques. These causes of variation were not analyzed because data were not available. Considering the results obtained from the analysis of the available data, it is reasonable to conclude that little additional information would be gained from further study.

Process Control

Evaluation of the results presented in this section provides insight into the main sources of variation at a particular plant. To reduce fluctuations in effluent quality, for example in Plant No. 2, sludge settleability must be improved. In Plant No. 17, flow equalization might be used to reduce the perturbations observed in the effluent quality; in Plant No. 41, improvement of unidentifiable variables, such as human factors, might reduce effluent quality variations.

In practical applications, regulation of wastewater temperature and influent flow and concentration is difficult; consequently, improvement of operational procedures and those parameters related to the solids separating facilities, appear to be the most promising and practical parameters available to process control.

TABLE 30. NUMBER OF TIMES THAT EACH VARIABLE WAS CONTRIBUTED SIGNIFICANTLY (GREATER THAN 5 PERCENT) TO EFFLUENT QUALITY VARIATIONS OF 21 PLANTS

Variable	BOD	SS
Wastewater Temperature	6	3
Time	4	3
Flow	1	6
SVI	2	5
MCRT	1	3
MLSS	1	2
Influent BOD	4	-
Influent SS	-	1

SECTION 7

ON SETTING STANDARDS FOR WATER QUALITY MANAGEMENT

To meet the objectives of the Federal Water Pollution Control Act Amendments of 1972, the United States Environmental Protection Agency (EPA) has set standards of effluent quality. The effluent limitations for secondary treatment as defined by EPA are as follows. (12)

> The arithmetic means of the secondary treatment effluent BOD and SS concentrations shall not exceed 30 g/m^3 in a period of 30 consecutive days, nor shall exceed 45 g/m^3 in a period of 7 consecutive days. In addition, the arithmetic mean of the concentration of BOD and SS remaining in the effluent over any 30-day period shall not exceed 15 percent of the arithmetic mean of the values in the influent (85 percent removal).

Two questions arise from this statement of secondary treatment standards. First, are the standards structured in the most appropriate manner and second, does the arithmetic average provide adequate information about effluent quality variations. These two questions are important because of the possible effects of process upset on receiving waters. Plant sizes have increased in recent years as communities have grown and regional systems have been introduced. As a result, the potential detrimental effects of process upsets have also increased. Thus, the manner in which standards are structured is increasingly important.

OBJECTIVE

The objective of this section is to examine and discuss the appropriateness of effluent discharge standards as they are currently established and to compare the current approach of using arithmetic mean to an alternative approach involving the use of the geometric mean concentration value (GM) for setting effluent discharge standards.

PROPOSAL OF GEOMETRIC MEAN AS A MEASURE OF CENTRAL TENDENCY OF EFFLUENT CONCENTRATION

In Section 2, the daily effluent BOD and SS concentration distributions were examined for activated sludge treatment plants. It was found that the lognormal distribution consistently fits best to the observed effluent BOD and SS data and that the lognormal distribution may be used to explain and/or predict effluent quality performance. Effluent quality data characteristically have a skewed distribution, and this fact must be recognized when stardards are being considered.

Measuring Central Tendency

The notion of central tendency is an attempt to describe "typical" or "average" behavior of a random quantity. The most widely used measures of central tendency are the mean ($\bar{X}$), the median (the middle data value or the average of the two middle values), and the mode (the value most frequently observed). A less commonly used measure of central tendency is the geometric mean which is denoted by

$$GM = \left(\prod_{i=1}^{n} x_i \right)^{1/n} \tag{24}$$

where: Π represents the product of all sample values, and X_i represents a sample of independent observations $X_1, X_2, \ldots, X_n$. It is this latter measure that is considered particularly appropriate for lognormal data and is recommended here for describing central tendency for activated sludge effluent concentrations.

For skewed or for heavy tailed distributions, the sample mean $\bar{X}$ has generally been found inappropriate because of the drastic impact of observations far from the bulk of the data (so-called "outliers"). Such observations occur with reasonably high probability when sampling from certain distributions, including the lognormal. The extent that outliers influence various measures of central tendency may be gleaned from Table 31.

Discussion

The lognormal distribution has a multiplicative reproductive property, that is, if $X_1, \ldots, X_n$ are independent, identically distributed lognormal variables, then the product $\prod_{i=1}^{n} X_i$ also has a lognormal distribution; concluded from the following theorem: (4) *If X_i (i=1, . . . , n) are independent lognormal variates with the same parameters (m_x, σ_x^2)*, their geometric mean $\left(\prod_{i=1}^{n} X_i \right)^{1/n}$ is lognormally distributed with parameters (m_x, σ_x^2/n). This fact has the important consequence that when observations have a given lognormal distribution, the exact distribution of the geometric mean is known, and the parameters are related in a simple way to the parameters of the parent distribution.

For a lognormal sample, the exact distribution of the mean does not belong to any known parametric family and the distribution of the median is too cumbersome to be of practical use. This fact alone provides strong support for the use of the geometric mean, since exact computations of probabilities like $P(GM \geq 45 \text{ g/m}^3)$ for a 7-day geometric mean are easily accomplished while similar computations for $\bar{X}$ or for the median are difficult and generally involve complex approximation methods.

TABLE 31. THE INFLUENCE OF OUTLIERS ON THREE MEASURES OF CENTRAL TENDENCY

Data set	Arithmetic mean, X	Median	Geometric mean, GM
1, 2, 3, 4, 5	3	3	2.61
1, 2, 3, 4, 50	12	3	4.13
1, 2, 3, 4, 500	102	3	6.54

TYPING GUIDE SHEET

The central limit theorem (27) might be invoked for probability computations involving $\bar{X}$, but the small sample size from which $\bar{X}$ is computed will sometimes preclude this possibility. This situation is illustrated by a simulation given in the appendix B. With this simulation it was shown that 7-day and 30-day averages of lognormal variables do not appear to be well described by a normal distribution.

The lognormal distribution often fits data for processes having an underlying multiplicative character, that is, where the observed variables are themselves products of more basic variables. In problems with a multiplicative character, such as effluent BOD or SS concentration, the geometric mean has a unique interpretation which adds to the motivation for its use as a measure of central tendency. This interpretation may be illustrated by the following example.

Consider a biological population that grows in each time period of a given length roughly in proportion to its own size. If its initial size is N, then its size in successive time periods can be written as: NX_1, NX_1X_2, . . . , $N(\prod_{i=1}^{n} X_i)$; where X_i represents the observed growth factor during the ith period. The growth factors themselves might be considered independent, identically distributed variables $X_1, \ldots, X_n$. The geometric mean of these variables is the "typical" or "average" growth factor in the sense that if this growth factor were experienced in each of the n time periods, the resulting population size would be precisely the population size observed. In other words, the geometric mean represents the value of a constant growth factor required to achieve the observed population size after n periods of growth.

An additional property of the geometric mean is that for large samples it is approximately lognormally distributed regardless of the distribution of the original data as long as the second moment of the distribution of LnX is finite. (8) Thus, even when the lognormal assumption for the original variable fails, it may be possible to make reliable probability computations for the geometric mean by treating it as a lognormal variable.

WATER QUALITY STANDARDS BASED ON GEOMETRIC MEAN

Use of the geometric mean of daily effluent quality data for setting standards in water quality management is recommended here. This recommendation is based on the fact that the geometric mean of lognormal data has known probabilistic behavior and a meaningful interpretation.

Use of the geometric mean instead of the arithmetic mean in setting and interpreting discharge standards would not necessarily require a change in the definition of secondary treatment or of the reference concentration values used. That is, the required values could be a 30-day and 7-day geometric mean concentration of 30 g/m^3 and 45 g/m^3, respectively. Because the inequality ($GM \leq \bar{X}$) holds for any data set from which these measures are computed, plants in violation under the proposed standards would also be in violation under the current standards. Clearly the converse is not true (as can be seen from Table 31) and thus use of the geometric mean values would result in somewhat less stringent discharge standards if no modifications of the standard values are made. This situation should not be a significant problem for three reasons. First, standard values now in use were developed on the basis of very limited information and may be changed if an appropriate rational is put forward. Second, many discharges are already operating under different (usually more stringent) standards. Third, a first step standard defined on a geometric mean basis can be set so that there is no effective change in the requirements. This is possible because of the amount of data that has now been collected and processed.

It should be noted that the idea of using geometric mean as a measure of central tendency is not unfamiliar in the field of biological treatment. Geometric mean has been used for many years as a measure of central tendency of coliform counts and as standards of this variable.

The relative insensitivity of the geometric mean to outliers in a data set guarantees that the measure will reflect the typical or average behavior of the process. There are circumstances in which an outlier is significant, and its presence should be noted. For example, if the value of effluent BOD in a secondary treatment plant is extremely high on a particular day, this may be considered a violation of the standard since by itself it may represent a significant amount of contamination. The standard in current use in some states (e.g., California) declares a violation if effluent BOD or SS on any single day exceeds an upper limit concentration. It is thus recommended that future standards be based on thresholds for single day, seven-day, and thirty-day geometric means. Since the value obtained from a single day may be viewed as the geometric mean of a sample of size one, the probabilistic features of the geometric mean discussed earlier may be used both for design and for monitoring considerations when dealing with data values taken one at a time.

SECTION 8

DISCUSSION

The primary objective of this study was to analyze statistically the effluent quality performance of a large number of activated sludge wastewater treatment plants. Such an analysis makes it possible to examine the natural variations that are observed in effluent quality and to introduce the concepts of reliability and stability in the design and operation of treatment processes.

Benefits derived from the study can be grouped into two areas dealing with 1) design and operation, and 2) setting of discharge requirements. Applications in design and operation include the introduction of stochastic concepts into the design process and the development of parameters for use in the design and operation of stable and reliable processes. Recommendations for setting discharge requirements include stochastic definitions for standards and the use of the geometric mean as opposed to arithmetic mean value in defining the central tendency of effluent BOD and SS concentration data.

FACTS

As shown in Sections 3 and 6, effluent BOD and SS concentrations from activated sludge processes and the variables affecting effluent quality in all plants are highly variable. The range of the variations depends both on internal and external factors, some of which cannot be controlled and are unavoidable.

It is generally known that the input loads, environmental conditions, and operational parameters are probabilistic in nature and are subject to random fluctuations in their magnitude, composition, and simultaneous occurrence. These variables all may affect the biological treatment of wastewater with subsequent effects on process performance and effluent quality. This variability should be recognized in process design and when evaluating effluent quality by discharge standards.

Examination of BOD and SS concentration distributions of the 43 activated sludge plants studied in this work has resulted in the following general findings. First, effluent BOD and SS concentration histograms are positively skewed. Second, in most cases, effluent concentration data are lognormally distributed. When the effluent concentration of a plant follows a lognormal distribution, there is always a possibility that effluent BOD and SS concentrations exceed any stated finite upper limit. Therefore, the lognormal distribution should be recognized in process design and in setting discharge standards.

In general, most wastewater treatment facilities are designed with limited information on the nature of the variations in input loads and environmental conditions. The usual procedure is that the flow records, if they exist, are analyzed and some peak flow rate is established, and an estimate is made of the corresponding BOD and SS mass loadings (with little or no rational basis for the decision other than experience). Design parameters such as mean cell residence time, sludge wastage, and sludge recycle rate are established on the basis of past experience and intuition.

Process design is usually based on empirical kinetic models in which steady state conditions are assumed, with kinetic coefficients obtained from the literature. Kinetic models currently available are based on mean values; thus, variations in performance cannot be predicted using these models. Seldom, if ever, is the stochastic nature of the input loading, environmental, and operational parameters incorporated into the process design.

Process design involves prediction of performance and decision on configuration and control strategy, too. Each of these facets of the design process involves some aspect of probability theory. Unfortunately, in only a small number of studies related to effluent quality and standards have probabilistic concepts been considered. Thus, techniques for incorporating information on performance variability into the design and standard setting process have not been implemented. Considering the ranges of process performance and variability found in this study and the fact that many processes are clearly not performing well in either respect, it would appear that inclusion of stochastic concepts into the design process would be beneficial.

STOCHASTIC BASED DESIGN AND STANDARDS

In view of the uncertainties in process design and operation, the performance of a given process will be uncertain to some degree. Therefore, design procedures should be based on a recognition of the uncertainties, that is, should be stochastic in character.

Inherent uncertainties can formally be recognized in the use of stochastic models, and thus, these models are closer to reality than deterministic models. (38) Moreover, stochatic models can be constructed in a manner which facilitates the incorporation of new information. Use of probabilistic design methods allows for process uncertainties and provides a means to quantify this uncertainty from available information.

Probabilistic considerations may be introduced both in design and in setting discharge requirements. Use of probabilistic methods in setting discharge standards precludes the rational setting of parameter values than can never be exceeded. For example, setting maximum allowable values of BOD or SS for any single sample has no meaning if a measure of random variation in these parameters exists. An excellent analogy, familiar to most Civil Engineers, is found in the return period of floods. Values of the 100, 200, or 1000-year floods are often predicted from data that include flood records of only a few years, but the larger floods are assumed to be possible.

Probabilistic procedures can be formulated in such a manner as to provide a probabilistic basis for design even with a deterministic standard. Given the natural variability of plant performance, it should be possible to determine what true average effluent quality must be maintained such that a standard can be met with some high (and predetermined) probability. In other words, the requirements should be rephrased to permit the parameter to exceed the designated value, say, one time in 3 years or once in 10 years. The required plant performance can then be calculated.

Design Procedure

In this part, the design of an activated sludge process, with respect to reliability and stability, is discussed.

Reliability--The designer of a treatment plant should first select an appropriate level of process performance. In most cases, a minimum reliability requirement should be established and, as discussed in Section 4 under Reliability Concept, a statistical-economic decision should be made. Plant cost, both capital and operating, rise with increasing performance requirements. On the other hand, lesser expectations in water quality or acceptance of a higher probability of violations will result in adverse effects on water quality downstream. Thus, the level of reliability and the effluent discharge standards directly affect plant economics.

After selection of a minimum reliability level for effluent quality, variations in the process should be estimated. The effluent variability may be described by the coefficient of variation. Selection of the appropriate design value of the coefficient of variation, V_x, must be based on reasonable expectations of performance. This coefficient may be estimated for each process type using the results reported in Table 16, Section 5.

The coefficient of reliability (COR) may be determined using Equation 20 or the data given in Table 5 or Figure 7. The COR relates mean constituent values to the standard and is calculated on a probability basis. This coefficient can then be used to design a treatment process that is expected to perform at a certain level of reliability by finding the required design effluent BOD or SS, M_x.

To design a process that will produce an effluent quality less than or equal to X_s, a stated $1-\alpha$ percent of the time when the coefficient of variation is V_x, a mean effluent concentration value equal to or less than COR(X_s) should be used. For example, if the effluent BOD is to be less than or equal to 30 g/m^3, at least 95 percent of the time, with a coefficient of variation of 0.7, the COR is 0.43 (Table 5) and it would be necessary to design the process to produce an average effluent BOD no greater than 13 g/m^3 (.43 x 30). As noted in Section 5, Table 16, activated sludge process can be expected to operate on the average at $V_x = 0.7$ for effluent BOD and $V_x = 0.84$ for effluent SS.

The design effluent BOD value is then used in kinetic models. It is important to note that this value should include both soluble and non-soluble BOD. Most kinetic modeling studies will result in prediction of very low soluble effluent BOD values. Then, the primary effluent BOD loading is due to suspended solids.

It is realized that procedures do not currently exist for design on the basis of effluent suspended solids concentration. Design values for suspended solids should be converted to equivalent BOD concentration values based on available conversion factors. (28)

Stability--Examination of the descriptive statistics of 43 treatment plants resulted in the conclusion that plants having a standard deviation of less than 10 g/m^3 for both effluent BOD and SS may statistically be considered as "stable" plants. Therefore, to produce a stable effluent quality, the process must be designed for a mean value equal to or less than $m_x = S_x/V_x$, where S_x is the standard deviation and is equal to or less than 10 g/m^3. The design mean values of effluent BOD and SS concentration with respect to stability are given in Table 16, Section 5.

Process Design Example--Assume the objective is to design a conventional activated sludge process to produce an effluent BOD that is equal to or less than 30 g/m^3, 95 percent of the time. The estimated design coefficient of variation is $V_x = 0.69$ for effluent BOD (Table 16). To produce a stable effluent quality, the design mean value for BOD should be equal to or less than 14.5 g/m^3. $(m_x = S_x/V_x = \leq 10/.69)$ To produce an effluent BOD equal to or less than 30 g/m^3, 95 percent of the time, the COR for this process is 0.43 (Table 5 at $V_x = .69$ and 95 percent reliability) and the corresponding design mean value is 12.9 g/m^3 (30 g/m^3 x .43) Therefore, this plant should be designed for an effluent mean BOD value equal to or less than 12.9 g/m^3. Exactly the same procedure may be followed to estimate the design mean SS value. Then, the SS concentration may be converted to its equivalent BOD value when the kinetic models are used in the design of an activated sludge process.

Operational Application

Application of the results of the study in plant operation is straighforward and is similar to design applications. The information given in Table 6 (or Figure 7) may be used to estimate the reliability of a treatment plant (i.e., the fraction of time that effluent concentration exceeds any standard value) if the mean and standard deviation for the plant is known. For instance, from Table 6 it can be seen that over a long period of time, about 15 percent [(1-.85) x 100] of the observed effluent SS values would exceed 30 g/m^3 in a complete-mixed wastewater treatment plant operating with a mean SS value of 18 g/m^3 $(m_x/X_s = 18/30 = 0.6)$ and standard deviation of 18 g/m^3 $(V_x = S_x/\bar{X} = 18/18 = 1.00)$. Since the standard deviation of this plant is greater than 10 g/m^3, the plant is classified as unstable, statistically. There is a chance that the maximum SS value will exceed 60 g/m^3.

STANDARDS

It is because of the lognormal distributional characteristics of effluent BOD and SS that the geometric mean is recommended for describing the central tendency of data. Present EPA standards for discharge effluent BOD and SS are based on arithmetic 7 and 30-day averages. As discussed in Section 6, these average values do not adequately represent the central tendency of a skewed distribution data set. On the other hand, by stating the standard in terms of the geometric mean values, one important consequence is that when observation have a given lognormal distribution, the exact distribution of the geometric mean is known (because of the multiplicative reproductive property of the lognormal distribution discussed earlier in Section 6). Further, the distribution parameters of the geometric mean values (mean and standard deviation) are related in a simple way to the parameters of the parent population distribution. Therefore, replacing the arithmetic 7 and 30-day mean value standards with the geometric 7 and 30-day mean values is a more accurate way of describing the central tendency of the data.

Reliability Estimation and the Use of the Geometric Mean

An advantage of stating standards in terms of the geometric mean as opposed to the arithmetic mean is that the reliability model presented in Section 3 may also be applied to the 7 and 30-day average standards. As mentioned in Section 6, the mean and standard deviation of the 7 and 30-day geometric mean values can be calculated easily if the mean and standard deviation of the parent population is known. This fact extends the use of the methodology presented for reliability estimation of single day values to the 7 and 30-day geometric means, since these variables are themselves lognormally distributed. Therefore, when standards are written in terms of achieved levels based on the geometric mean, it will be possible to determine precise requirements for a treatment plant. This is illustrated in the following example.

Example—In a plant operating with a mean BOD value $m_{BOD} = 18\ g/m^3$ and standard deviation $\sigma_{BOD} = 12.6\ g/m^3$, the mean and standard deviation of the 30-day geometric mean (GM) values can be estimated using Equations 12 and 13 and Theorem 1[a], as follows: The mean and standard deviation of the logarithm of BOD values are 2.69 and 0.63, respectively (Equations 12 and 13).

$$\sigma^2_{LnBOD} = \text{Ln}\left(\frac{\sigma^2_{BOD}}{m^2_{BOD}} + 1\right) \qquad (12)$$

$$m_{LnBOD} = \text{Ln}\ m_{BOD} - \frac{1}{2}\ \sigma^2_{LnBOD} \qquad (13)$$

---

[a] Theorem 1. If X_i, $(i=1, \ldots\ldots, n)$ are independent lognormal variates with the same mean and variance (m_y and σ^2_y), their geometric mean $(\prod_{i=1}^{n} y_i)^{1/n}$ is also lognormally distributed with parameters $(m_y, \sigma^2_y/n)$. The interpretation of this theorem in this example is as follows: Let m_{LnX} and σ^2_{LnX} represent the mean and variance of each variate Ln X_i. Denote the distribution of X_i by LN $(m_{LnX}, \sigma^2_{LnX})$. Then, the distribution of the geometric mean $(\prod_{i=1}^{n} X_i)^{1/n}$ is LN $(m_{LnX}, \sigma^2_{LnX/n})$. That is, Ln $(\prod_{i=1}^{n} X_i)^{1/n}$ is normally distributed with mean m_{LnX} and variance $\sigma^2_{LnX/n}$.

Therefore:

$$\sigma_{LnBOD} = 0.63 \quad \text{and} \quad m_{LnBOD} = 2.69$$

Based on Theorem 1, the 30-day geometric mean values are lognormally distributed. The logarithm of the geometric mean has a mean equal to 2.69 and a standard deviation equal to 0.115.

$$m_{LnGM} = 2.69$$

$$\sigma_{LnGM} = 0.63/\sqrt{30} = 0.115$$

The values of the mean and standard deviation for the original, 30-day geometric mean BOD values are 14.83 g/m^3 and 1.71 g/m^3, respectively, using Equations 12 and 13.

$$\sigma^2_{GM} = m^2_{GM} \left[\exp\left(\sigma^2_{LnGM}\right) - 1\right] \qquad (12)$$

$$L_n\, m_{GM} = m_{LnGM} + \frac{1}{2}\, \sigma^2_{LnGM} \qquad (13)$$

$$m_{GM} = 14.83$$

$$\sigma_{GM} = 1.71$$

The coefficient of variation is thus equal to

$$V_{GM} = \frac{\sigma_{GM}}{m_{GM}} = 0.115$$

and the reliability level for this plant will be greater than 99.9 percent based on a standard level of X_s = 30 g/m^3 (Equation 21 or Figure 7).

$$Z_{1-\alpha} = - \frac{Ln\left[\frac{m_{GM}}{X_s}\left(V^2_{GM} + 1\right)^{-½}\right]}{\left[Ln\left(V^2_{GM} + 1\right)\right]^{½}} = 5.95 \qquad (21)$$

Reliability > 99.9%

In other words, the chance that in a long run operation, the 30-day geometric mean values of the effluent exceeds the 30 g/m^3 standard, is very small.

The same procedure is applied in design. For example, the design mean value for a conventional sludge process that is to produce an effluent with a 30-day geometric mean value equal to or less than 30 g/m^3, 95 percent of the time, can be determined as follows: The coefficient of variation for effluent BOD is estimated to be equal to $V_{BOD} = 0.69$ (Table 16). The coefficient of variation of 30-day geometric mean values is equal to 0.114, using Equations 12 and 13 and Theorem 1^{a}.

$$\sigma^2_{LnBOD} = \text{Ln}\,(V^2_{BOD} + 1) = 0.389 \qquad (12)$$

Using Theorem 1, parameters of the logarithm of the 30-day geometric mean are:

$$\sigma^2_{LnGM} = 0.389/30 = 0.013$$

$$m_{LnGM} = \text{Ln}\; m_{GM} - 1/2\,(\sigma^2_{LnGM})$$

and also $$m_{LnGM} = \text{Ln}\; m_{BOD} - 1/2\,(\sigma^2_{LnBOD})$$

Therefore: $$m_{GM} = m_{BOD}/1.207$$

and $$\sigma^2_{GM} = m^2_{GM}\,[\exp \sigma^2_{LnGM} - 1]$$

Solving the above equations results in,

$$V_{GM} = \frac{\sigma_{GM}}{m_{GM}} = 0.114$$

From Equation 20 (or Figure 7), the value of the coefficient of reliability (COR) at 95 percent reliability (e.g., $Z_{1-\alpha} = 1.645$) is equal to 0.83, and the design mean effluent BOD value to meet 30 g/m^3 standard will be 0.83 x 30 = 24.9 g/m^3. If the level of reliability for this plant is increased to 99.9 pecent, (e.g., $Z_{1-\alpha} = 3.090$), the design value should be reduced to 21.2 g/m^3.

It should be noted that the same procedure presented here is applied to 7-day geometric mean values and also can be applied to suspended solids concentration as well.

Process Control

The correlation and regression analysis (Section 5) resulted in the conclusion that a consistent relationship between effluent concentration and influent loads, biological, and operational variables did not exist. Generally, a single variable or group of variables cannot be used to explain the variability to a significant extent in all plants. Also, in defining plant performance variations, the contribution and influence of any variable was different for each plant. This situation may be a result of design, operation, or both. The fact is that the deterministic models now in use that are based on steady state assumption do not include stochastic variability of any type construction and have been demonstrated to be unsuitable for application to non-steady state system. (36) For example, variations in influent variables, temperature (which may cause changes in current patterns in clarifiers) and dissolved oxygen concentration are not modeled well.

The overall results of variability study are too dispersed to generalize the sources and extent of variability of process performance. However, the correlation or regression analysis discussed in Section 5 may be applied to data from eacn individual treatment plant to determine the principal factors responsible for effluent quality variations. This is especially important when a plant is not operating satisfactorily. By identifying the main sources of the variation, corrective actions may be taken to reduce variability and to improve the stability of the plant.

An additional comment is appropriate. At the present time, for most plants it is required that influent, effluent, and operational variables be measured and reported on a daily basis. Therefore, an extensive set of operating data are available for most plants. Often this data collection process becomes useless and is done only for the sake of reporting. Yet it should be recognized that valuable information is contained in the data. For instance, by applying the methods presented in this paper, these data can be used to explain the causes of the variation in effluent quality. Finding the major cause of the variations may help the operator to improve plant performance.

REFERENCES

1. Adams, B. J., and Gemmell, R. S., "Performance of Regionally Related Wastewater Treatment Plants." **Jour. Water Poll. Control Fed.**, 45, 2088 (1973).

2. Adams, B. J., and Gemmell, R. S., "Mean Estimate Deficiencies in Water Quality Studies." **Jour. Hydraulics Div., Proc. Amer. Soc. Civil Engr.**, 101, HY7, 989 (1975).

3. Adams, C. E., Jr., "Techniques of Monitoring for Variability Control." **Progress in Water Technology**, 8, 201 (1976).

4. Aitchison, J., and Brown, J. A. C., **The Lognormal Distribution**, London, Cambridge University Press (1957).

5. Andrews, J. F., "Review Paper, Dynamic Models, and Control Strategies for Wastewater Treatment Processes." **Water Research**, 8, 261 (1974).

6. Ang, A. H-S, and Ellingwood, B. R., "Critical Analysis of Reliability Principles Relative to Design," **Statistics and Probability in Civil Engr.** Proceedings of the First International Conference on Applications of Statistic and Probability to Soil and Structural Engineering, Hong Kong, (Sept. 1971).

7. Ang, A. H-S., "Structural Risk Analysis and Reliability Based Design." **Jour. Structural Div., Proc. Amer. Soc. Civil Engr.**, 99, ST9, 1891 (1973).

8. Benjamin, J. R., and Cornell, A., **Probability Statistics, and Decision for Civil Engineers**, McGraw-Hill Book Company (1970).

9. Berthouex, P. M., "Some Historical Statistics Related to Future Standards." **Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr.**, 100, EE2, 423 (1974).

10. Chiang, C. H., "Process Stability of Activated Sludge Processes." **Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr.**, 103, EE2, 259 (1977).

11. Chorafas, D. N., **Statistical Processes and Reliability Engineering**, Van Nostrand Company, Ch. 39, 339 (1960).

12. **Code of Federal Regulations**, Title 40, Protection of Environment, Parts 100 to 339, p. 158 (1976).

13. Dean, R. B. and Forsythe, S. L., "Estimating the Reliability of Advanced Waste Treatment." Part I, **Water and Sewage Works**, 87 (June 1976).

TYPING GUIDE SHEE

14. Dean, R. B., and Forsythe, S. L., "Estimating the Reliability of Advanced Waste Treatment." Part II, **Water and Sewage Works,** 57 (July 1976).

15. Downing, A. L., "Variability of the Quality of Effluent from Wastewater Treatment Processes and its Control." **Progress in Water Technology,** 8, 189 (1976).

16. Eckenfelder, W. W., and Roth, J. A., "Effluent Variability as Related to Developing Performance Standards." **Progress in Water Technology,** 8, 1 (1976).

17. **Federal Water Pollution Control Act Amendments of 1972,** P.L. 92-500, S.2770 (October 18, 1972).

18. Ford, D. L., "Factors Affecting Variability from Wastewater Treatment Plants." **Progress in Water Technology,** 8, 91 (1976).

19. Forde, L., Niku, S., and Schroeder, E. D., "Discussion, Process Stability of Activated Sludge Processes." **Jour. Envr. Eng. Div., Proc. Amer. Soc., Civil Engr.,** 104, EE1, 178 (1978).

20. Forsyth, C. H., "Proposal of a Coefficient of Stability." **Jour. Amer. Statistical Association,** N.S. 27, 173 (1932).

21. Garrett, M. T., Jr., "Effluent Variability in Domestic Wastewater Treatment Plants." **Progress in Water Technology,** 8, 183 (1976).

22. Hann, R. W., et al., Evaluation of Factors Affecting Discharge Quality Variation. For the Texas Water Quality Board. Environmental Engineering Division, Civil Engineering Department, Texas AπM University, Austin, Texas (1972).

23. Hovey, W. H., Schroeder, E. D., and Tchobanoglous, G., **"Optimal Size of Regional Wastewater Treatment Plants."** California Water Resources Center, Contribution α161 (1977).

24. Krishnan, P., and Gaudy, A. F., Jr., "Response of Activated Sludge to Quantitative Shock Loading." **Jour. Water Poll. Control Fed.,** 48, 906 (1976).

25. Lawrence, A. W., and McCarty, P. L., "Unified Basis for Biological Treatment Design and Operation." **Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr.,** 96, SA3, (1970).

26. McKinney, R. E., "Mathematics of Complete-Mixing Activated Sludge." **Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr.,** 88, SA3 (1962).

27. Mendenhall, W., and Scheaffer, R. L., **Mathematical Statistics with Applications,** Duxbury Press (1973).

28. Metcalf and Eddy, Inc. (Revised by Tchobanoglous, G.), **Wastewater Engineering, Treatment, Disposal, Reuse,** Second Edition, McGraw-Hill Book Company (1979).

29. Neter, J. and Wasserman, W., **Applied Linear Statistical Models.** Richard Irwin Inc. (1974).

30. Nie, N. H., et al., **Statistical Package for the Social Sciences, SPSS.** Second Edition, McGraw-Hill Book Company (1975).

31. Niku, S., Schroeder, E. D., and Samaniego, F. J., "Performance of Activated Sludge Processes and Reliability Based Design." Accepted for Publication in **Jour. Water Poll. Control Fed.,** 51, 2841, (1979).

32. Novotny, V., "Response of First Order Biological Treatment Processes to Time Variable Inputs." **Progress in Water Technology,** 8, 15 (1976).

33. Popel, H. J., "A Concept for Realistic Effluent Standards." **Progress in Water Technology,** 8, 69, (1976).

34. Schroeder, E. D., **Water and Wastewater Treatment.** McGraw-Hill Book Company, New York (1977).

35. Schroeder, E. D., and Niku, S., "Biological Wastewater Treatment Plant Performance; Reliability, Variability and Stability." Interim Report to U.S. Environmental Protection Agency (April 1978).

36. Selna, M. W., and Schroeder, E. D., "Response of Activated Sludge Processes to Organic Transients - Kinetics." **Jour. Water. Poll. Control Fed.,** 50, 944 (1978).

37. Sewage Treatment at Military Installations. Report of the Subcommittee on Sewage Treatment of the Committee on Sanitary Engineering, National Research Council, Division of Medical Science, Washington, D. C., **Sewage Works Journal,** 18, 5, 791 (September 1946).

38. Shah, H. C., "The Rational Probabilistic Code Format." **ACI Journal,** 690 (1969).

39. Sherrard, J. H., and Lawrence, A. W., "Response of Activated Sludge to Step Increase in Loading." **Jour. Water Poll. Control. Fed.,** 47, 1848 (1975).

40. Singh, A., "How Reliable is the Factor of Safety in Foundation Engineering?" **Statistics and Probability in Civil Engineering. Proceedings of the First International Conference on Applications of Statistics and Probability to Soil and Structural Engineering,** Hong Kong, 390 (1971).

41. Singh, K. P., and Stall, J. B., "Variability of Wastewater Effluent Flow." **Progress in Water Technology,** 8, 237 (1976).

42. **Standard Methods for the Examination of Water and Wastewater,** 14th Edition, APHA . AWWA . WPCF (1976).

43. **Study of Discharge Requirements and Effluent Variability.** Sacramento Regional County Sanitation District, Sacramento, California (1975).

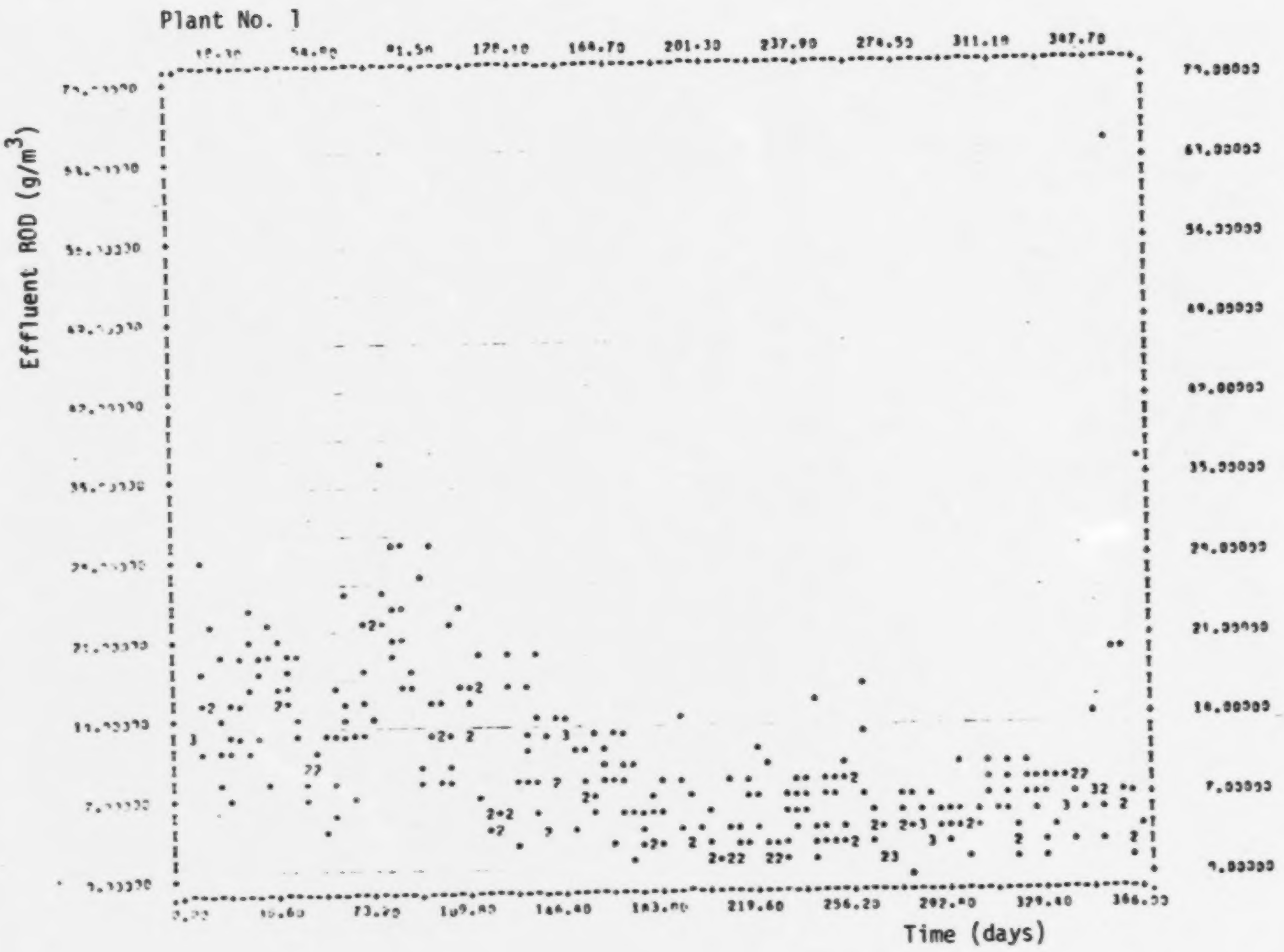
44. Thomann, R. V., "Variability of Waste Treatment Plant Performance." **Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr.,** 96, SA3, 819 (1970).

45. Townshend, A. R., "Statistical Analysis of the Effluent Quality of Biological Sewage Treatment Processes." **Proceedings the Third Canadian Symposium on Water Pollution Research,** Toronto, Canada, 272 (1968).

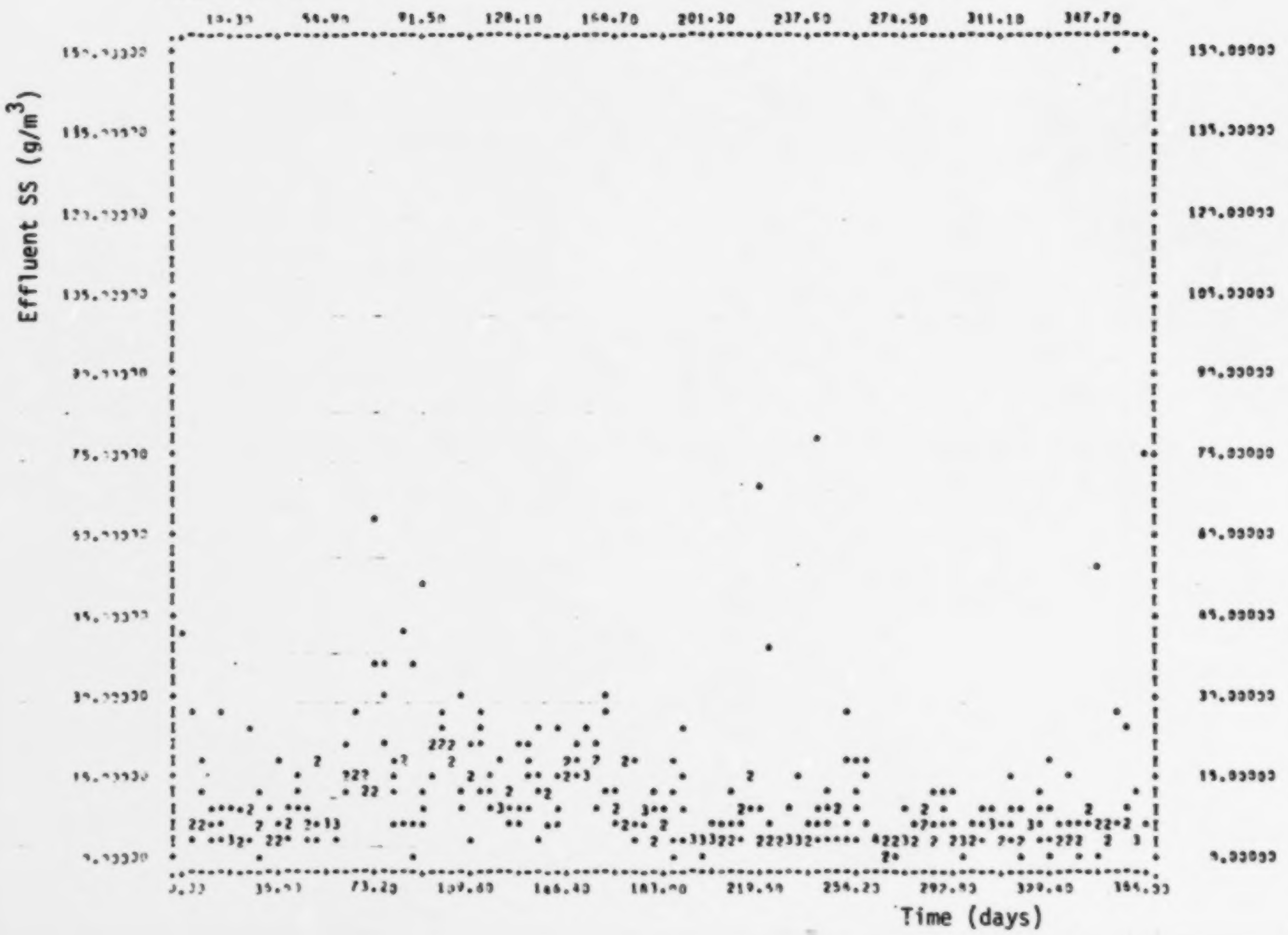
APPENDIX A

SCATTERGRAMS OF DAILY EFFLUENT BOD AND SS DATA

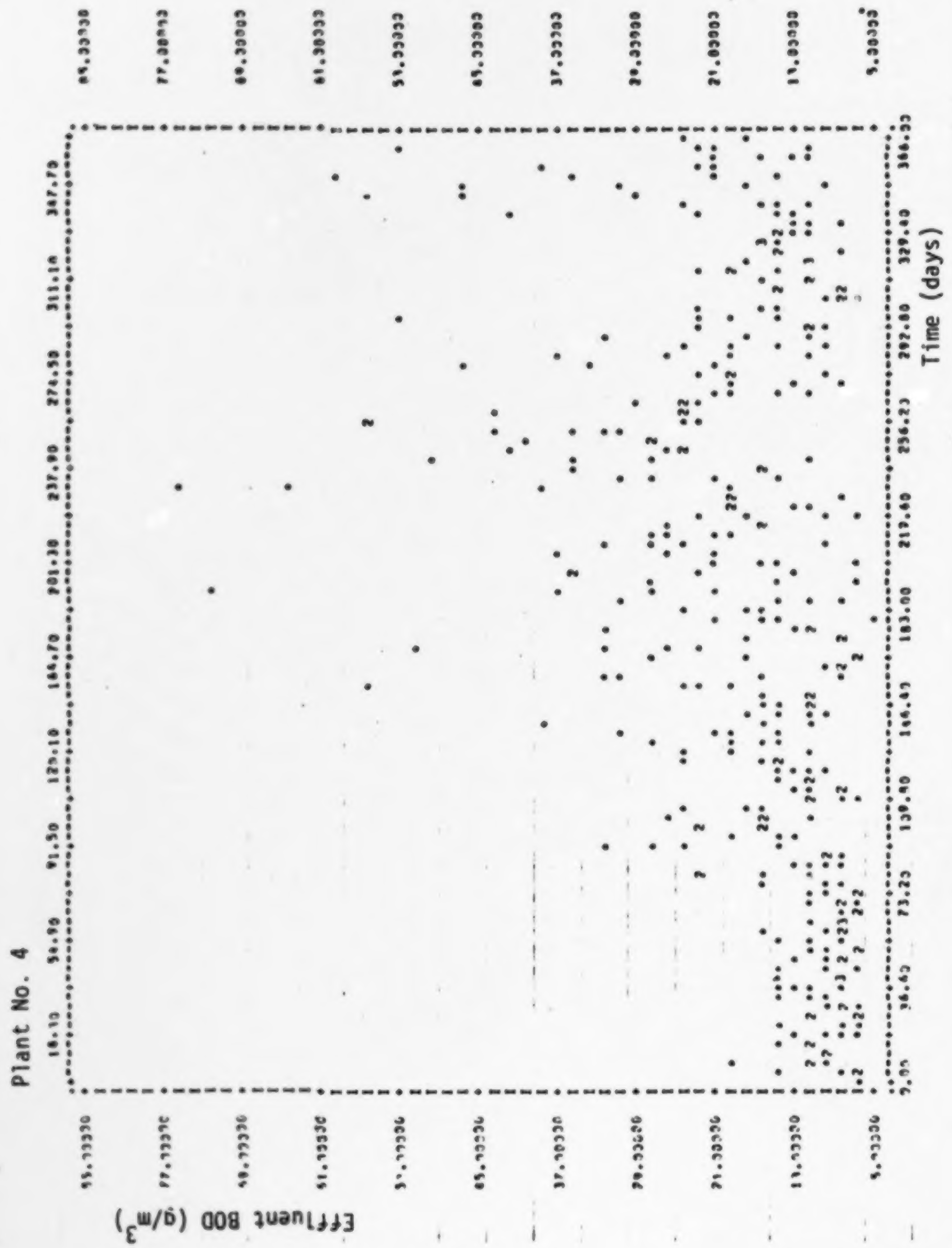
Selected time-series plots of effluent BOD and SS concentrations are presented in Appendix A. These plots are computer printouts of a two-dimensional plot of effluent concentrations versus time. The statistical package for the Social Sciences, SPSS (30), was used for plotting these scattergrams.

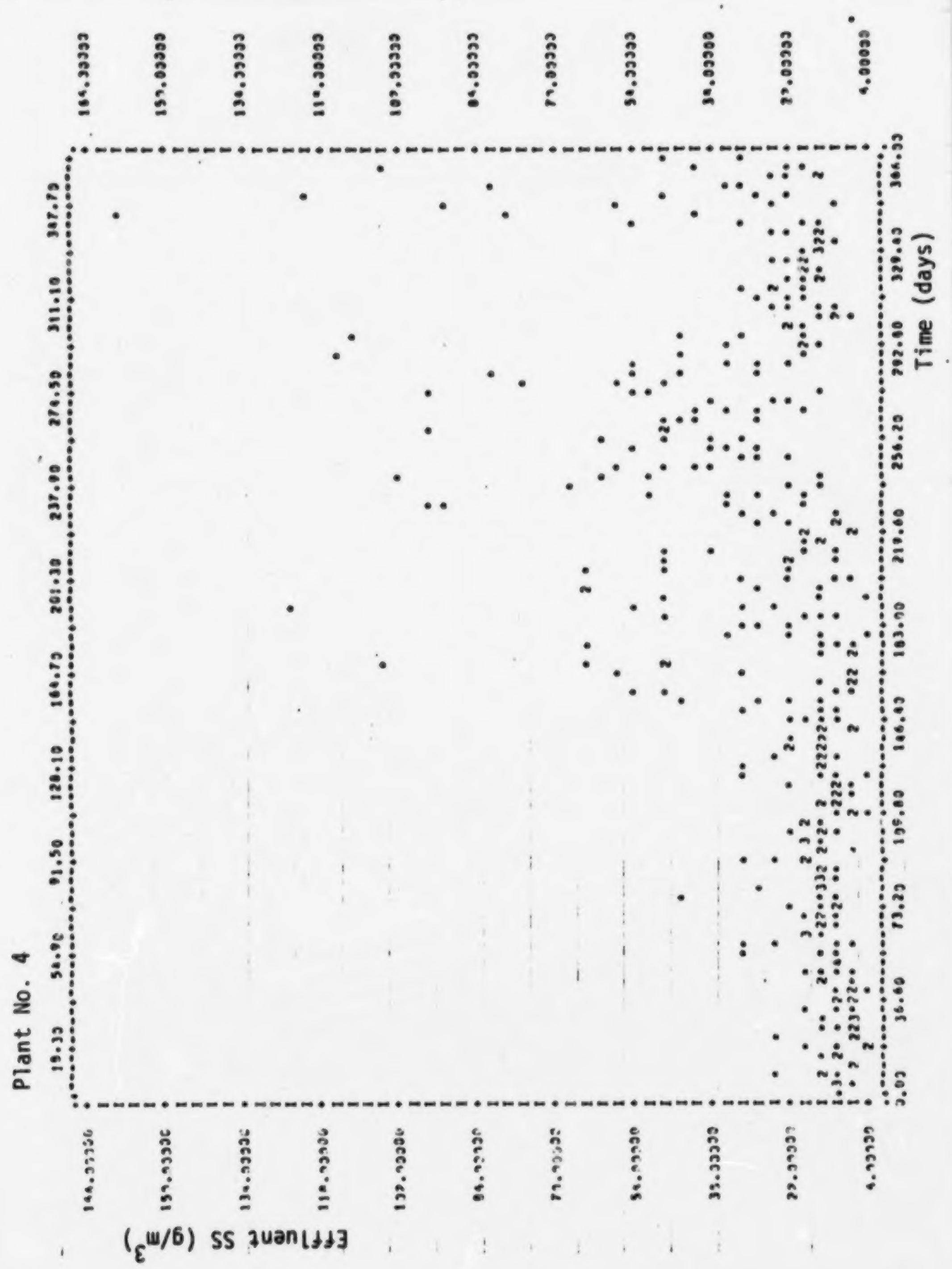

Plant No. 1
Effluent BOD (g/m³)
Time (days)
18.30
54.90
91.50
128.10
164.70
201.30
237.90
274.50
311.10
347.70
0.00
36.60
73.20
109.80
146.40
183.00
219.60
256.20
292.80
329.40
366.00
70.00000
63.00000
56.00000
49.00000
42.00000
35.00000
28.00000
21.00000
14.00000
7.00000
0.00000

Plant No. 1

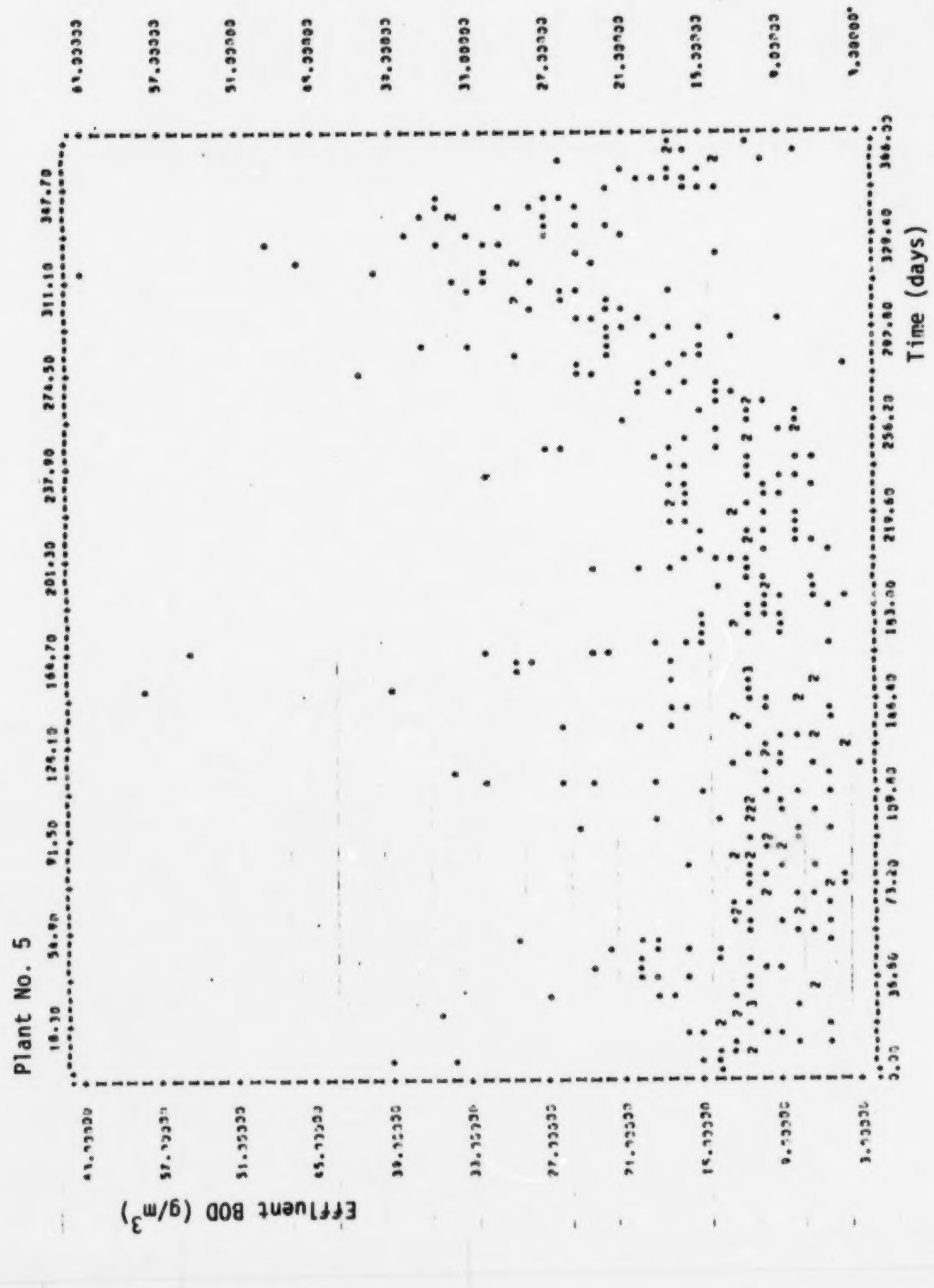


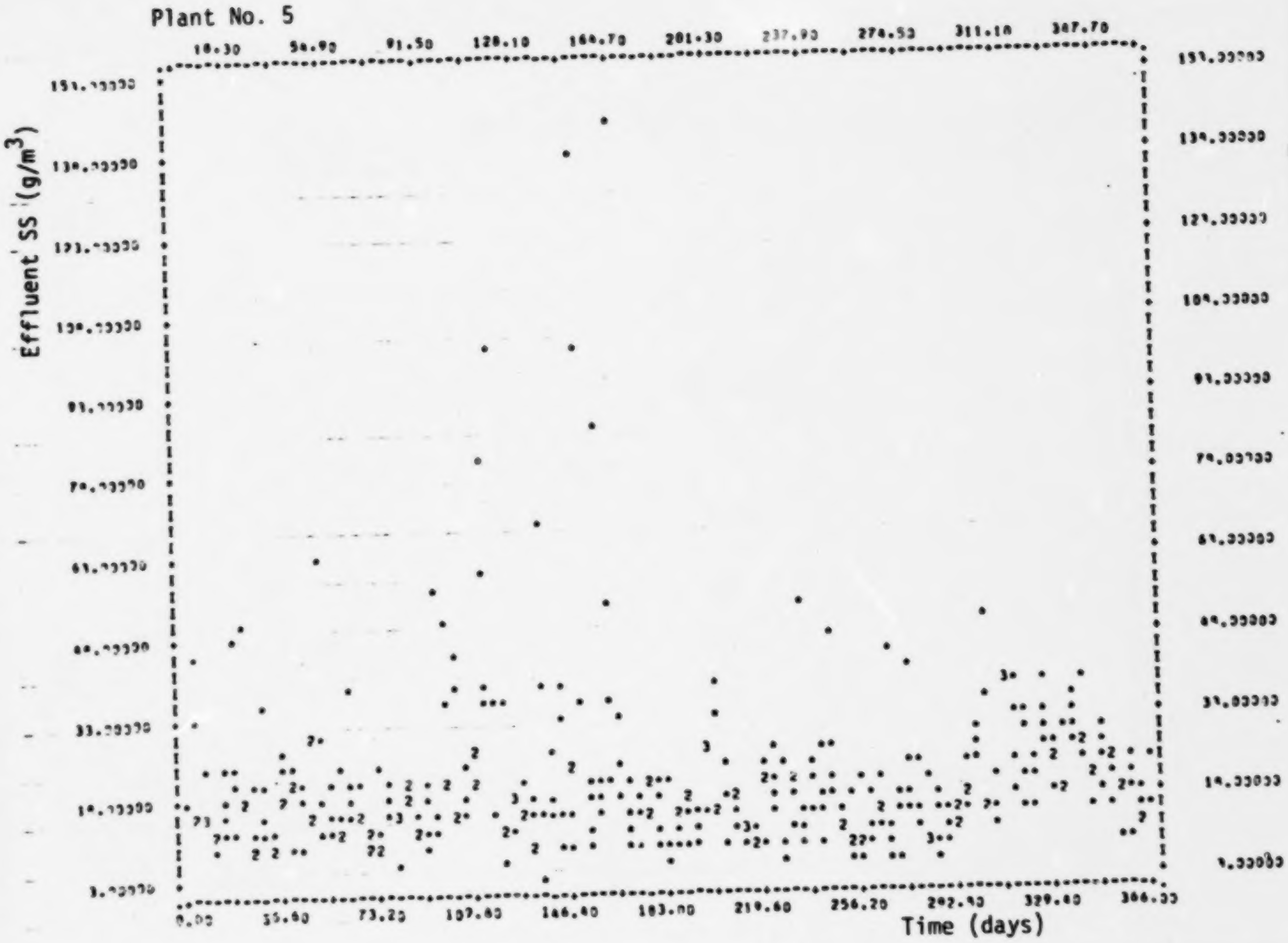
Effluent SS (g/m^3)
Time (days)
18.30
54.90
91.50
128.10
164.70
201.30
237.90
274.50
311.10
347.70
0.00
36.60
73.20
109.80
146.40
183.00
219.60
256.20
292.80
329.40
366.00
150.00000
135.00000
120.00000
105.00000
90.00000
75.00000
60.00000
45.00000
30.00000
15.00000
0.00000

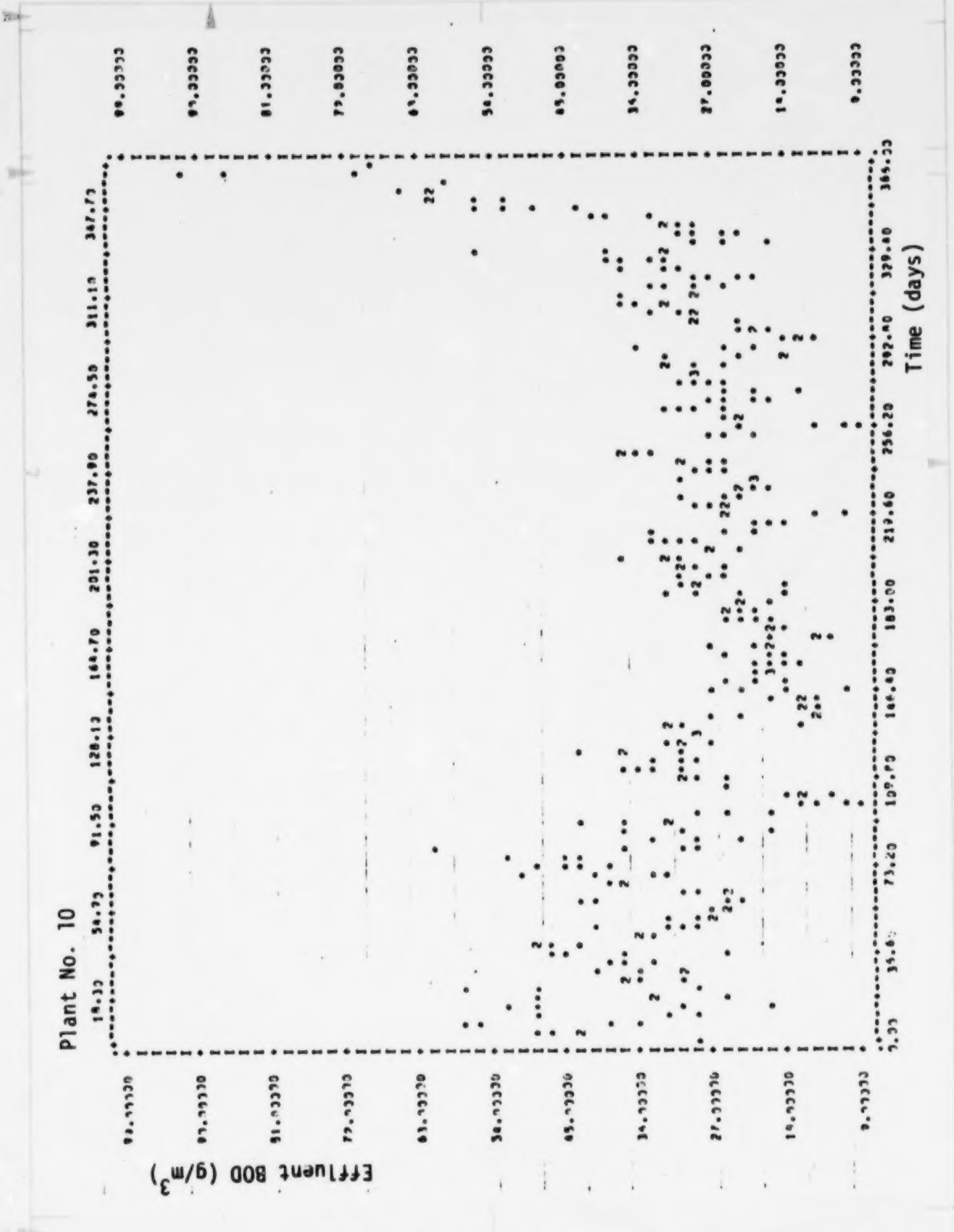

Plant No. 4
Effluent BOD (g/m^3)
Time (days)
85.00000
77.00000
69.00000
61.00000
53.00000
45.00000
37.00000
29.00000
21.00000
13.00000
5.00000
0.00
36.60
73.20
109.80
146.40
183.00
219.60
256.20
292.80
329.40
366.00
18.30
54.90
91.50
128.10
164.70
201.30
237.90
274.50
311.10
347.70



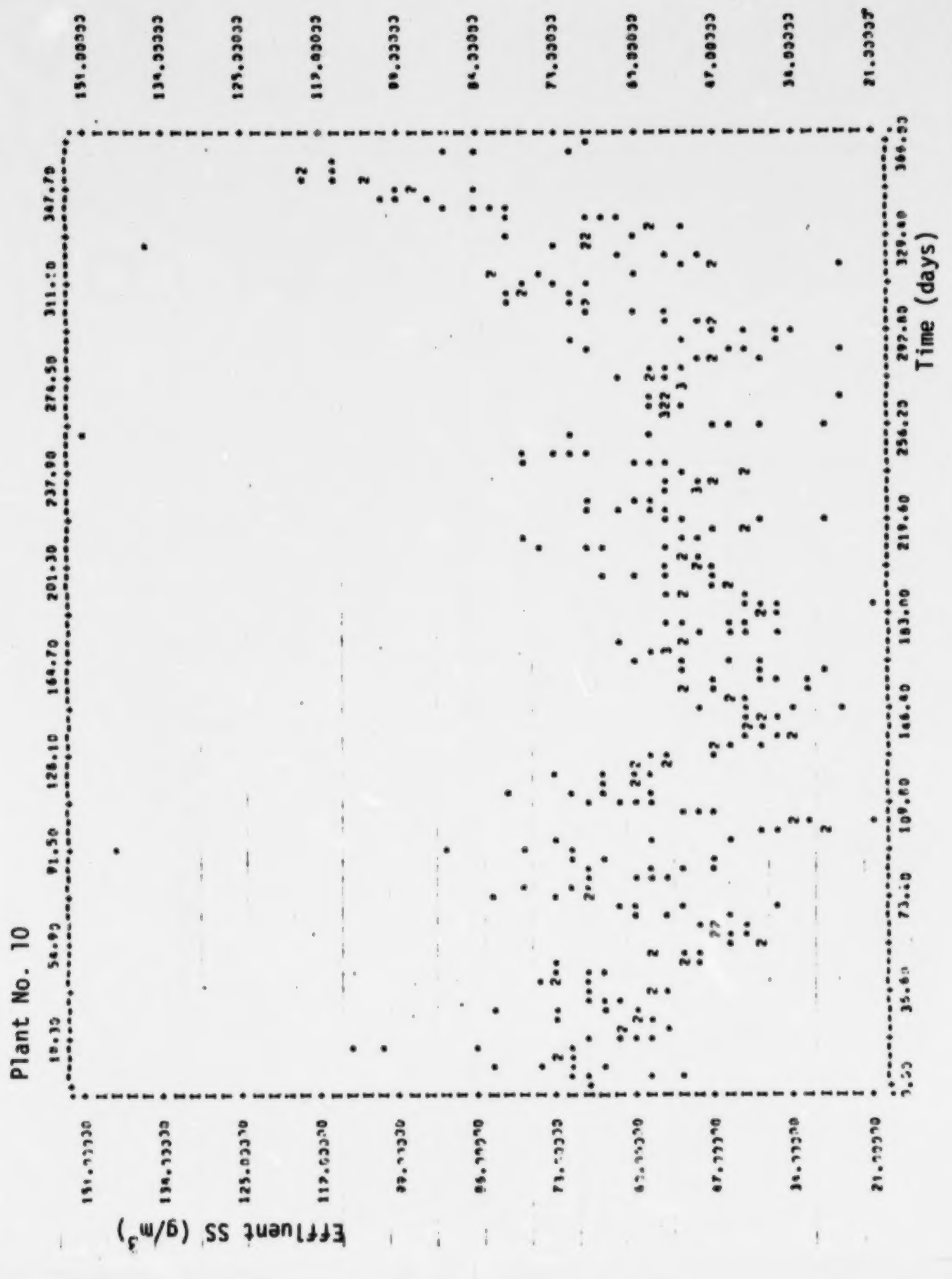
EPA Form 2230-4 (1-80)
TYPING GUIDE SHEET


Plant No. 5
Effluent BOD (g/m^3)
Time (days)
63.00000
57.00000
51.00000
45.00000
39.00000
33.00000
27.00000
21.00000
15.00000
9.00000
3.00000
0.00
18.30
36.60
54.90
73.20
91.50
109.80
128.10
146.40
164.70
183.00
201.30
219.60
237.90
256.20
274.50
292.80
311.10
329.40
347.70
366.00

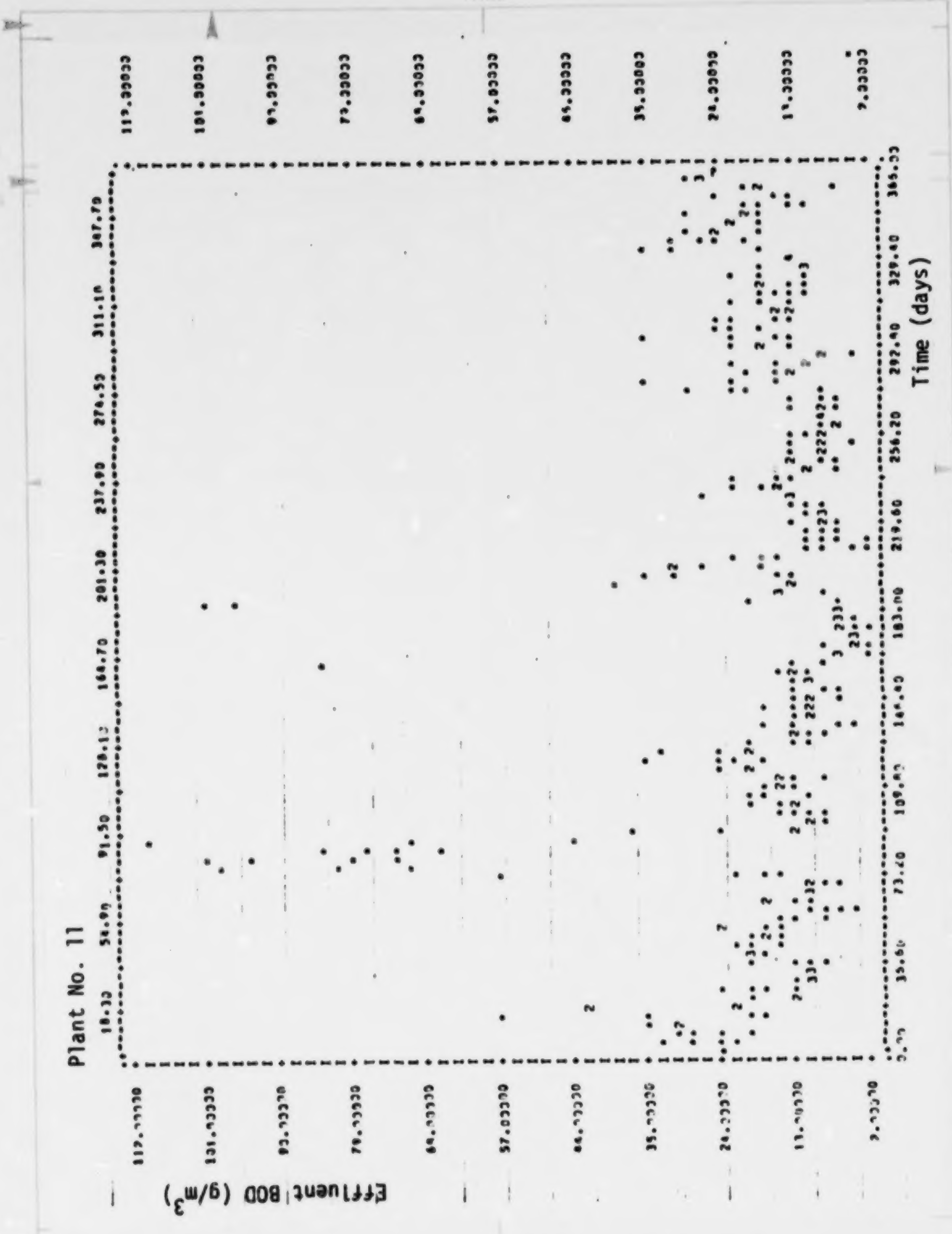

Plant No. 5
Effluent SS (g/m^3)
Time (days)
0.00
36.60
73.20
109.80
146.40
183.00
219.60
256.20
292.80
329.40
366.00
18.30
54.90
91.50
128.10
164.70
201.30
237.90
274.50
311.10
347.70
151.00000
136.00000
121.00000
106.00000
91.00000
76.00000
61.00000
46.00000
31.00000
16.00000
1.00000



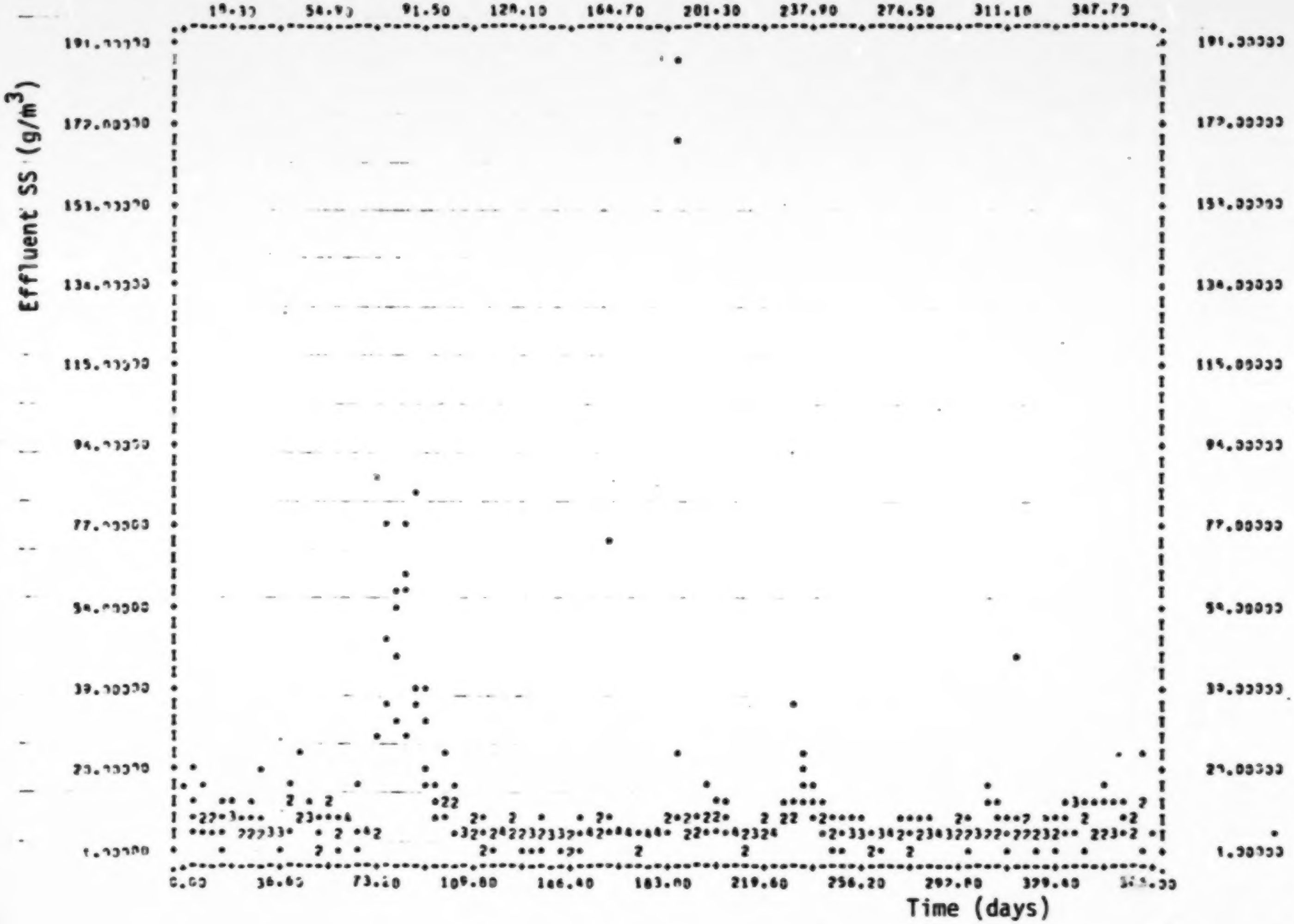
TYPING GUIDE SHEET



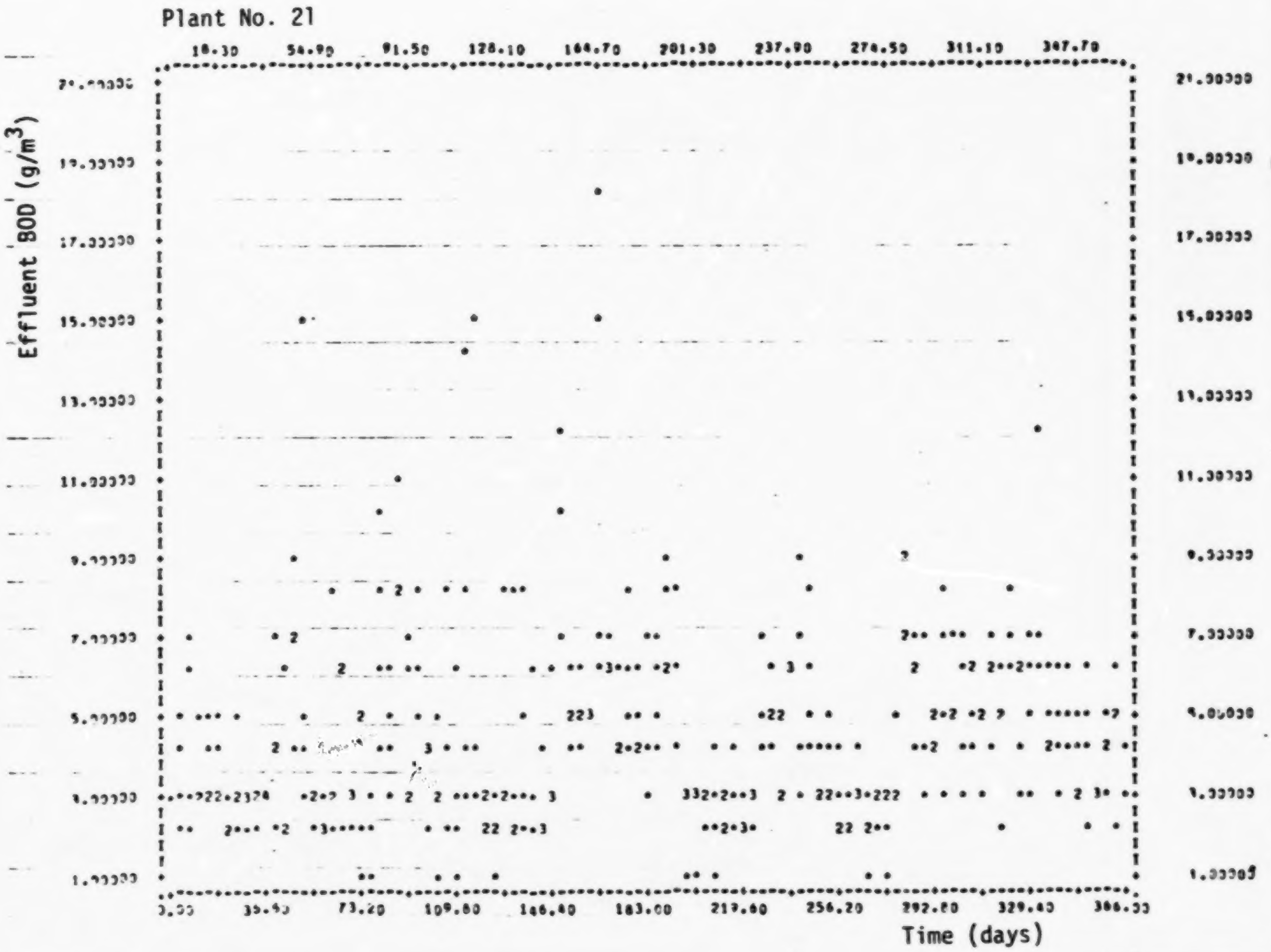
EPA Form 2350-4 (4-80)
(PREVIOUSLY LIN. EPA FORM 567)
TYPING GUIDE SHEET

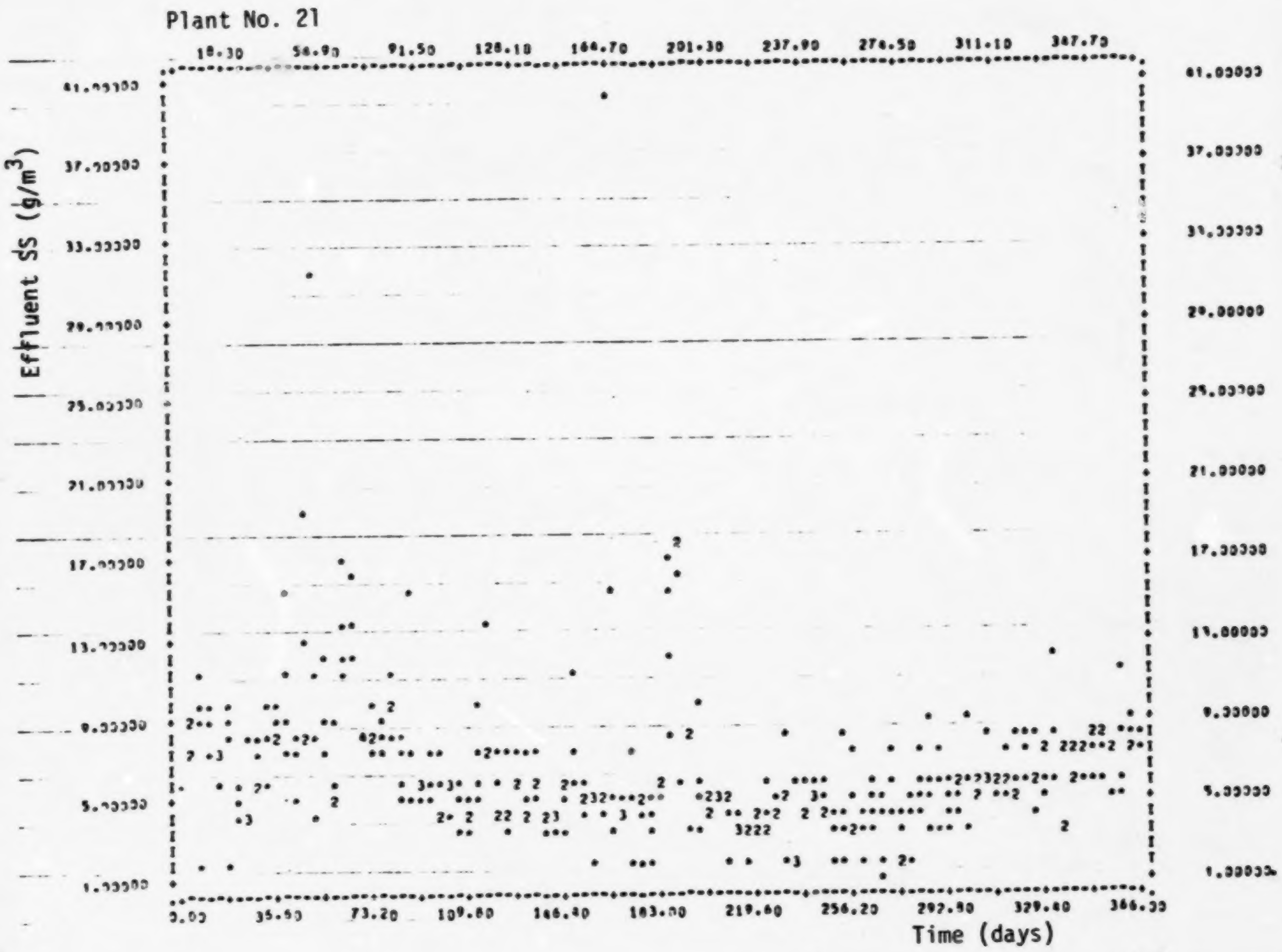

Plant No. 11
Effluent BOD (g/m3)
Time (days)
112.00000
101.00000
90.00000
79.00000
68.00000
57.00000
46.00000
35.00000
24.00000
13.00000
2.00000
0.00
18.30
36.60
54.90
73.20
91.50
109.80
128.10
146.40
164.70
183.00
201.30
219.60
237.90
256.20
274.50
292.80
311.10
329.40
347.70
366.00

Plant No. 11



TYPING GUIDE SHEET

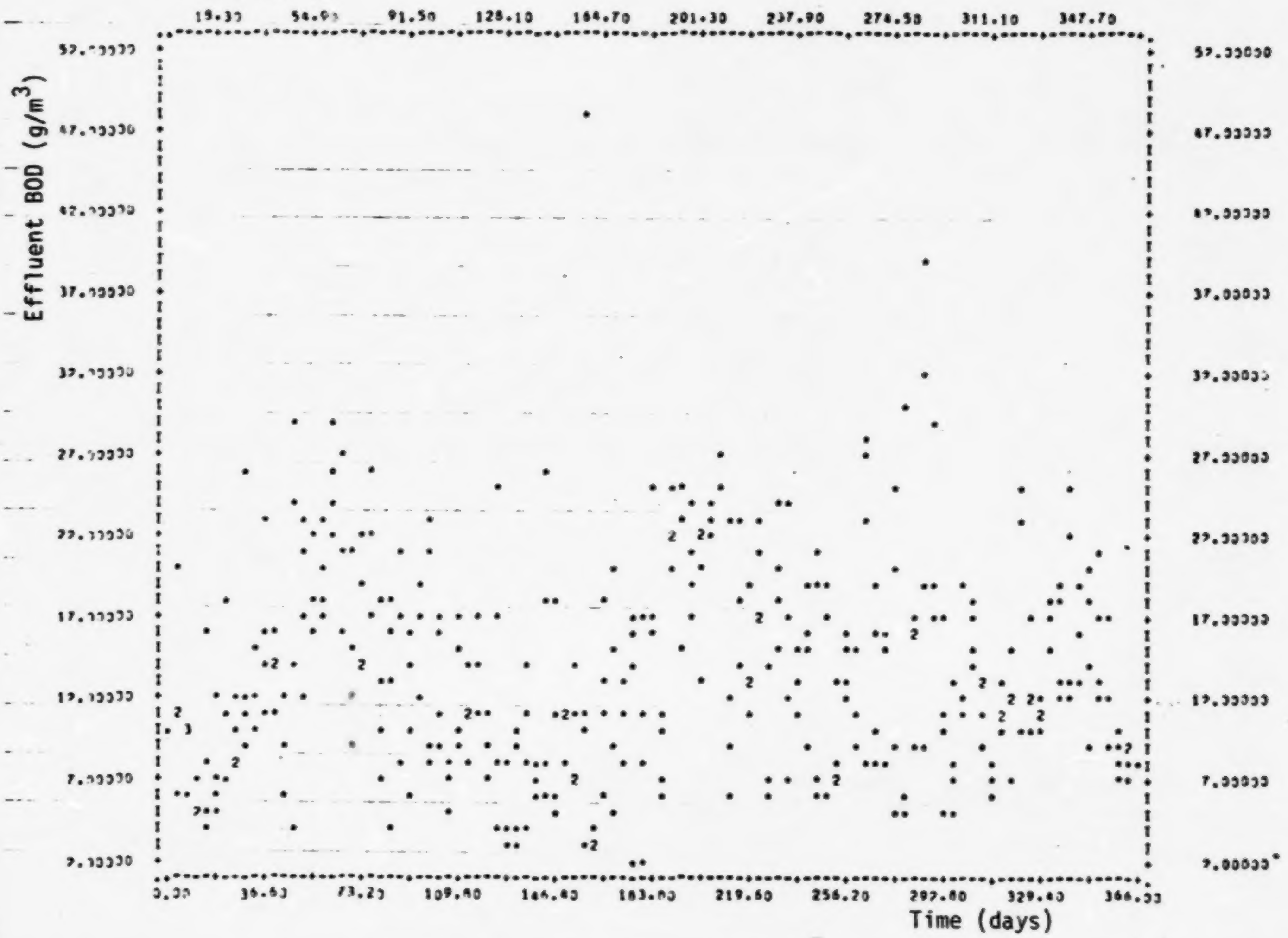

Plant No. 21
Effluent BOD (g/m^3)
Time (days)
18.30
54.90
91.50
128.10
164.70
201.30
237.90
274.50
311.10
347.70
21.00000
19.00000
17.00000
15.00000
13.00000
11.00000
9.00000
7.00000
5.00000
3.00000
1.00000
0.00
36.60
73.20
109.80
146.40
183.00
219.60
256.20
292.80
329.40
366.00



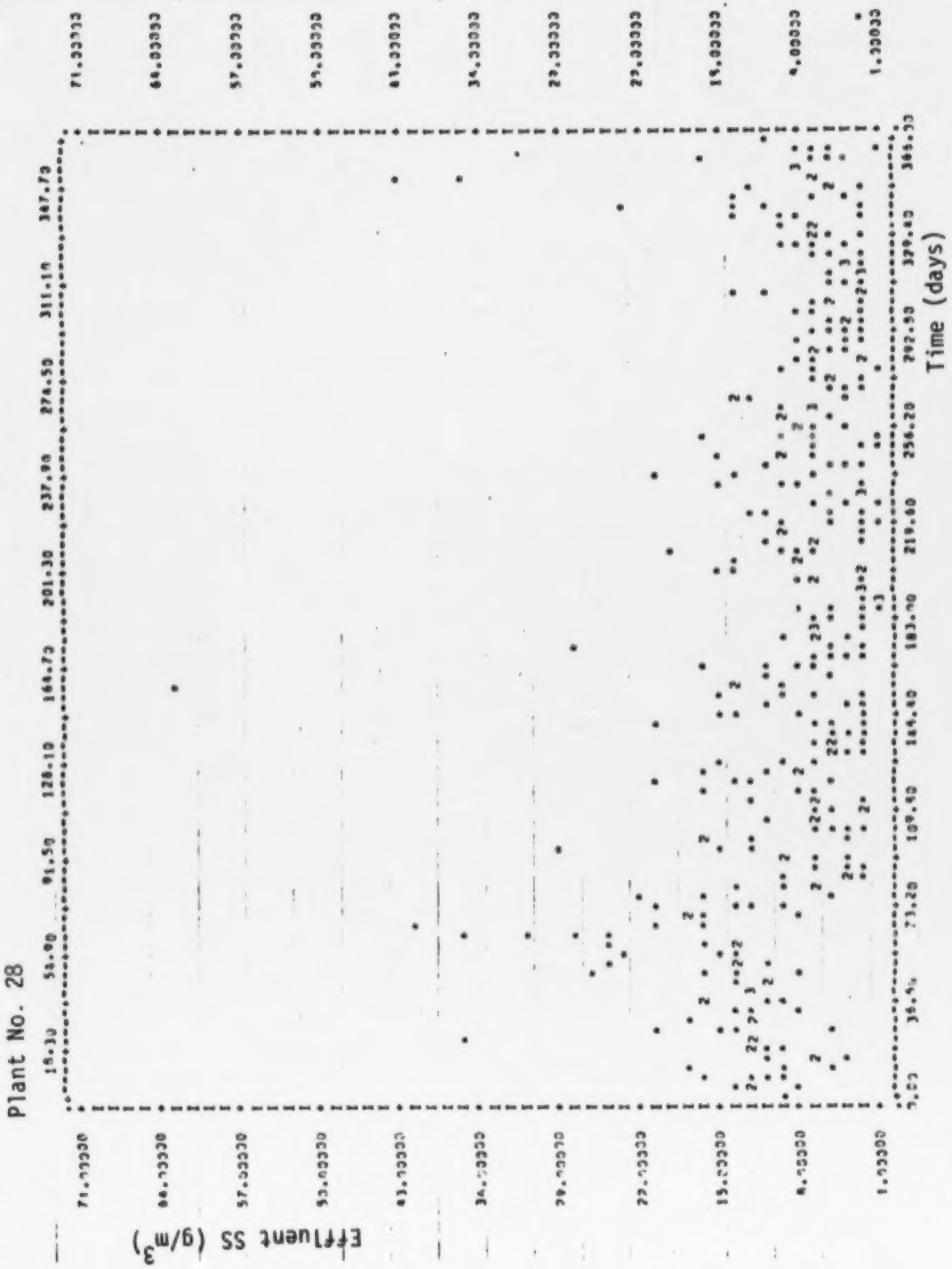
EPA Form 2300-6 (4-80)

TYPING GUIDE SHEET

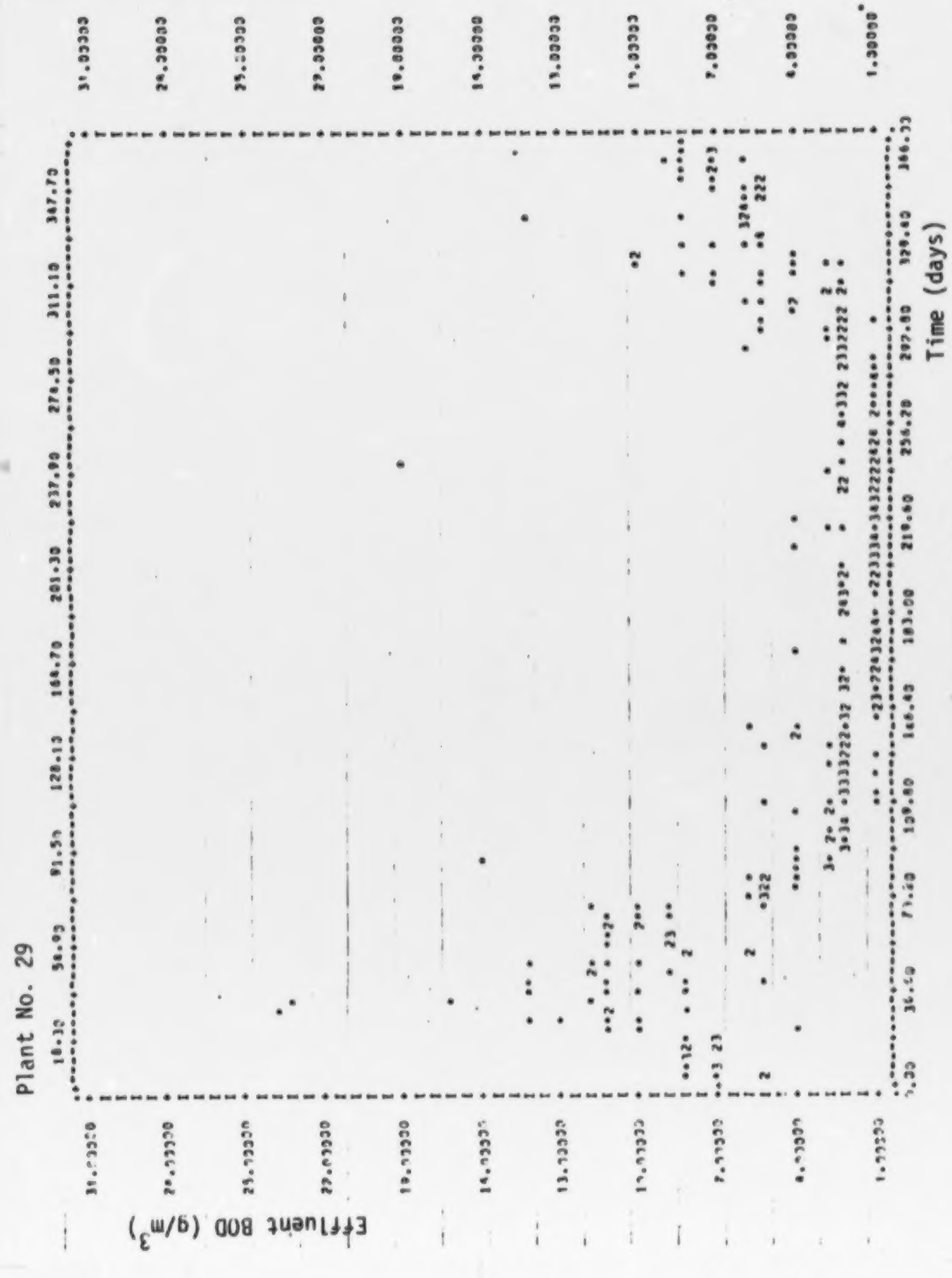
Plant No. 28



TYPING GUIDE SHEET

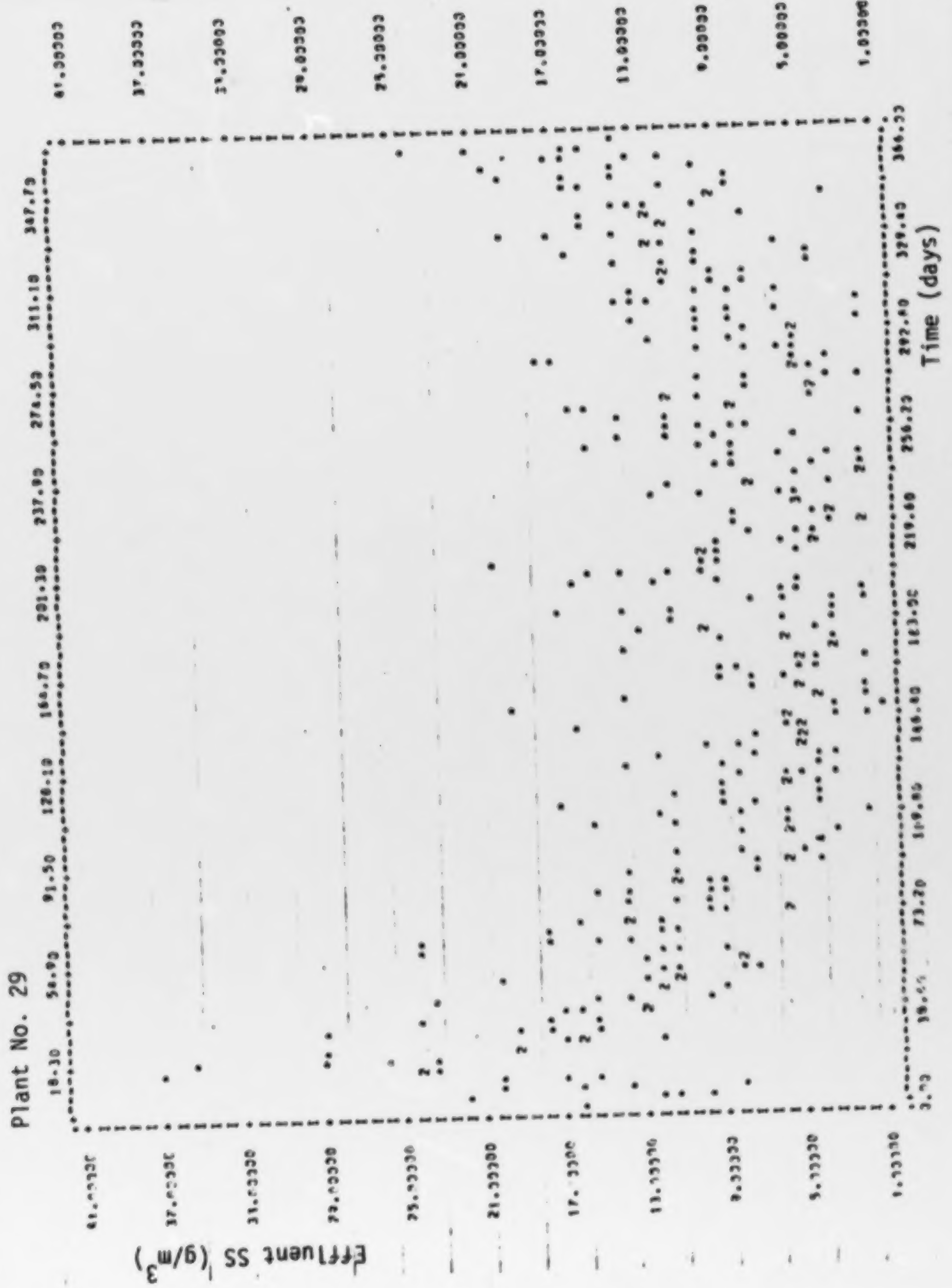


EPA Form 2350-4 (9-80)
TYPING GUIDE SHEET



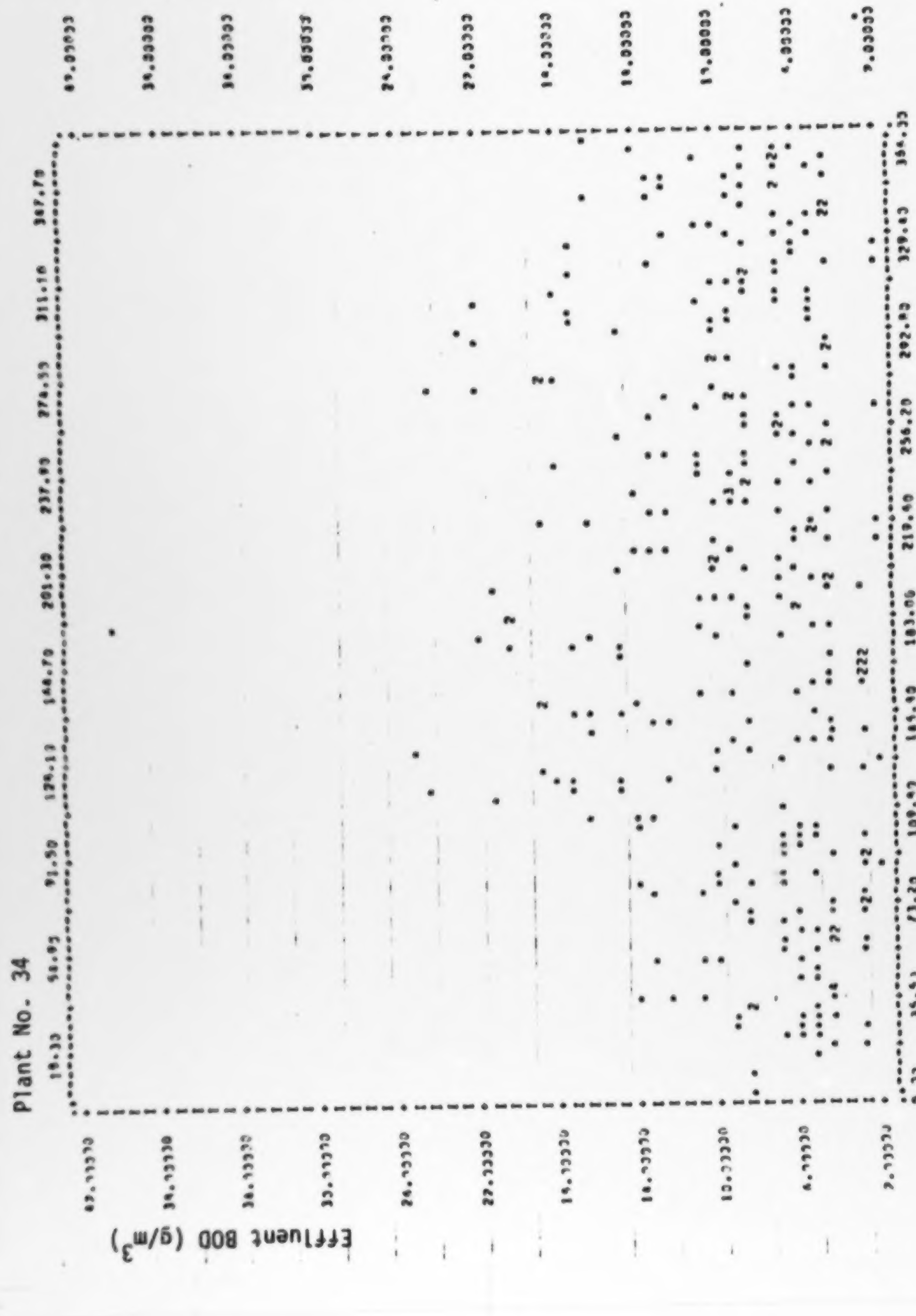
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE

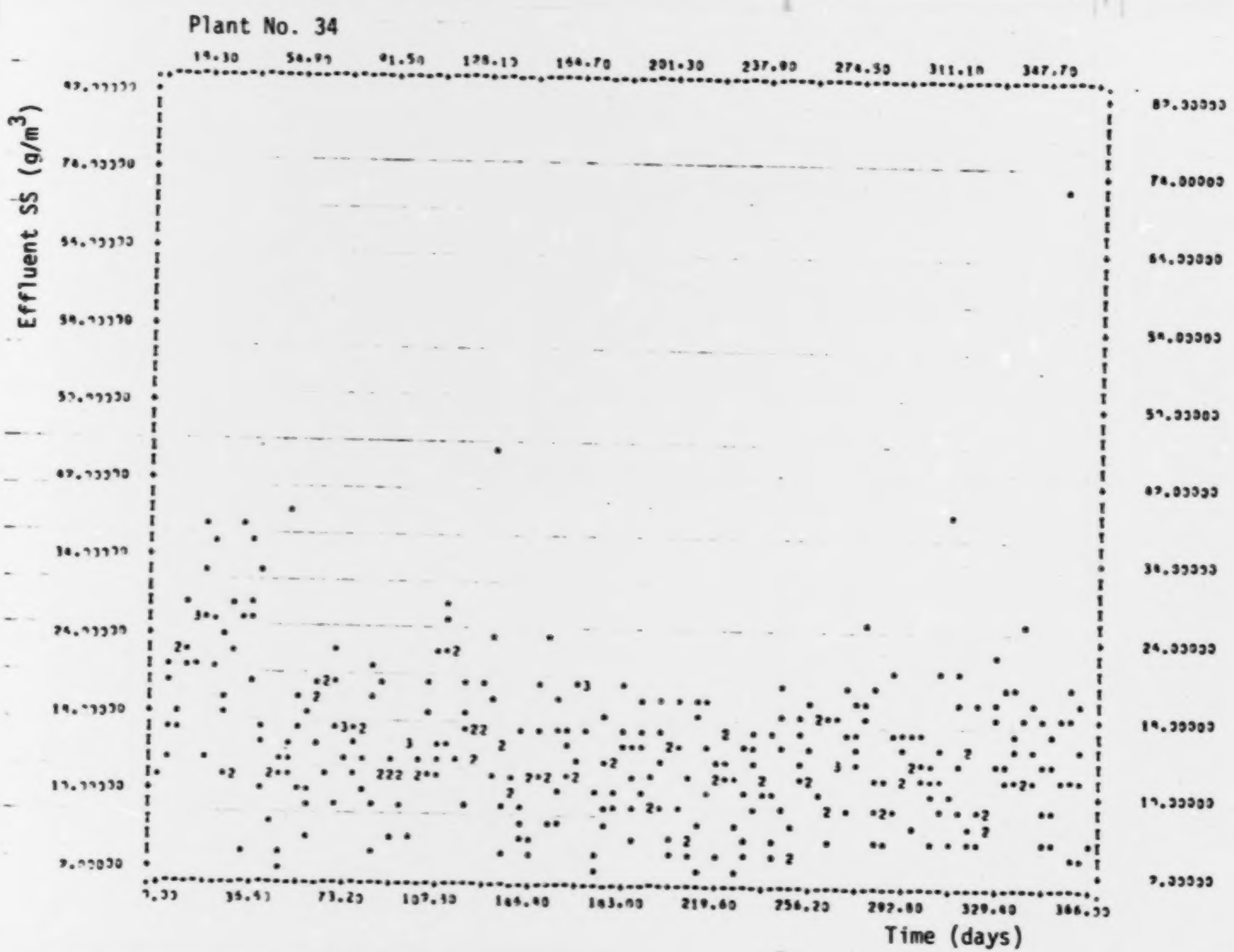


EPA Form 2200-4 (1-80)
TYPING GUIDE SHEET

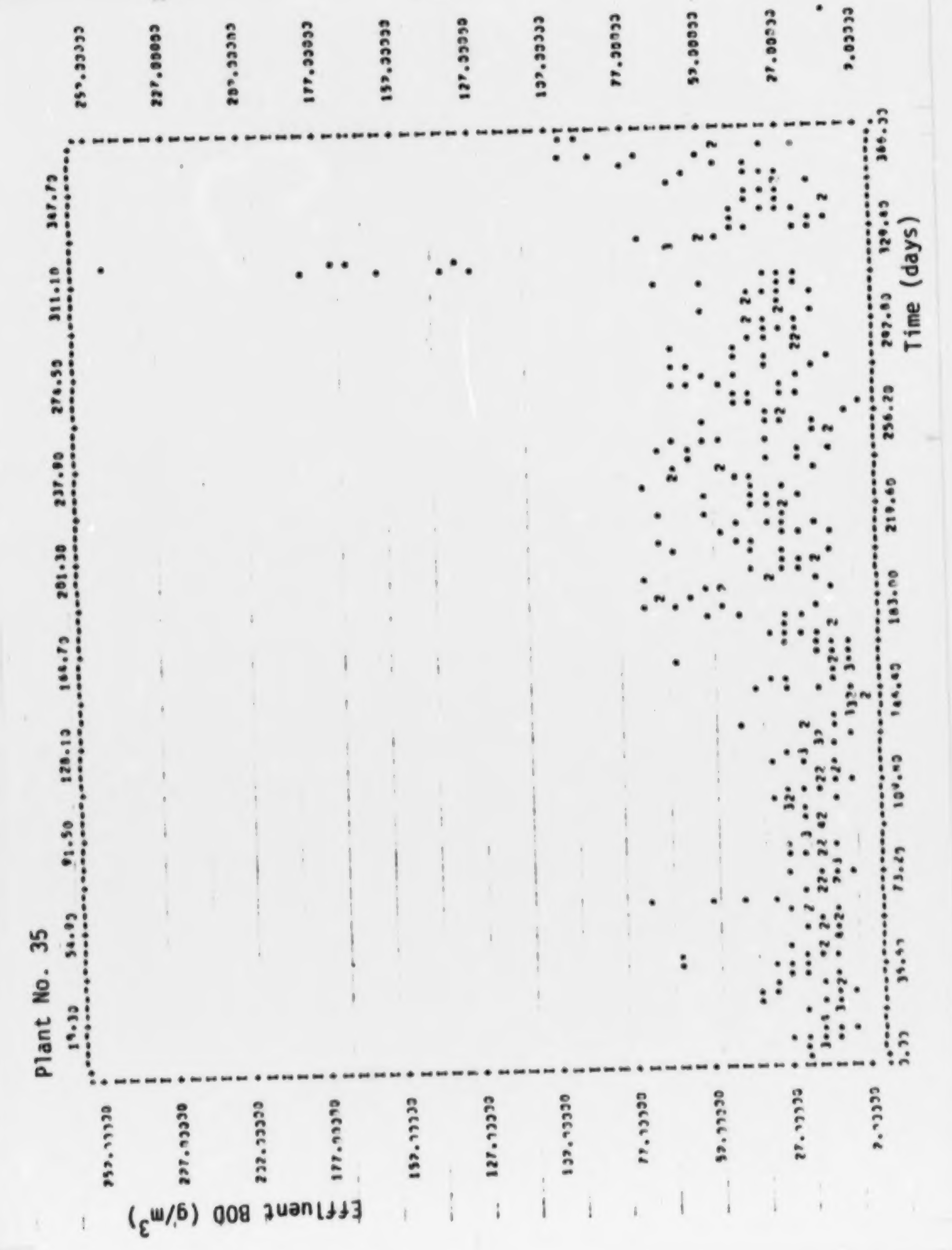
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA



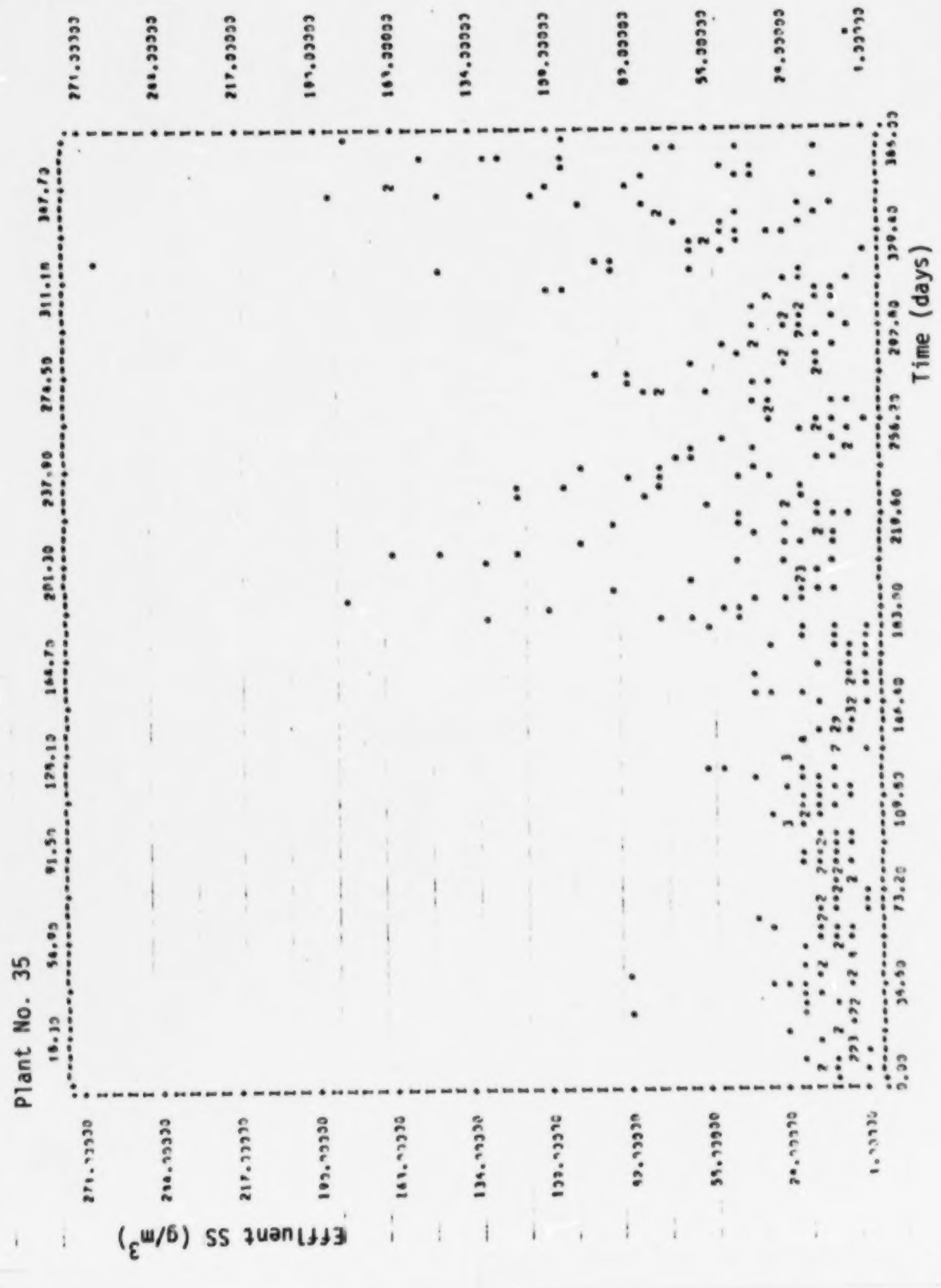
TYPING GUIDE SHEET

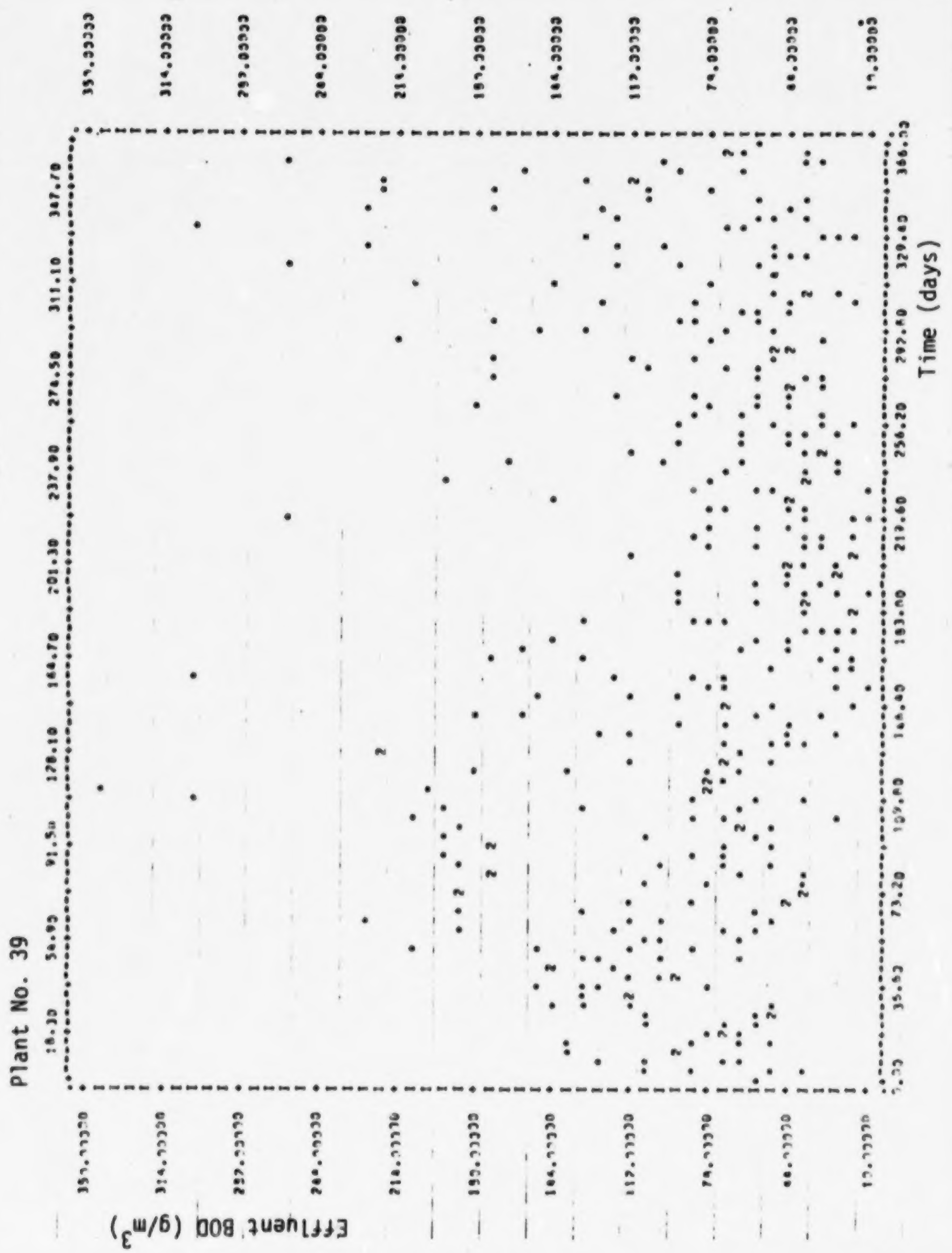


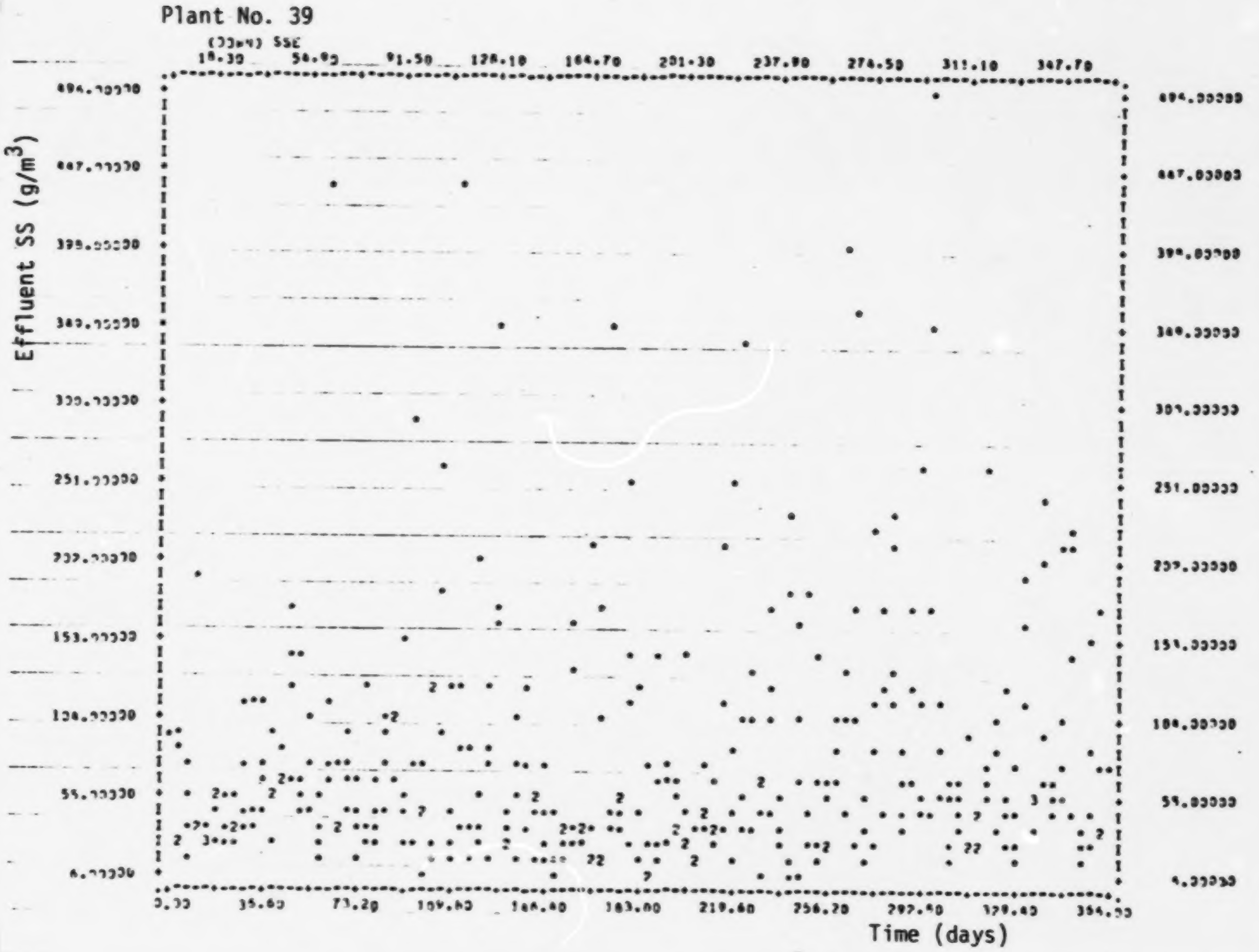
TYPING GUIDE SHEET



BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE
EPA Form 2200-4 (3-80)
TYPING GUIDE SHEET


Plant No. 35
Effluent SS (g/m^3)
Time (days)
0.00
18.30
36.60
54.90
73.20
91.50
109.80
128.10
146.40
164.70
183.00
201.30
219.60
237.90
256.20
274.50
292.80
311.10
329.40
347.70
366.00
1.00000
28.00000
55.00000
82.00000
109.00000
136.00000
163.00000
190.00000
217.00000
244.00000
271.00000


Plant No. 39
Effluent BOD (g/m^3)
Time (days)
12.00000
46.00000
79.00000
112.00000
146.00000
179.00000
213.00000
246.00000
280.00000
314.00000
347.00000
0.00
18.30
36.60
54.90
73.20
91.50
109.80
128.10
146.40
164.70
183.00
201.30
219.60
237.90
256.20
274.50
292.80
311.10
329.40
347.70
366.00


Plant No. 39
(DAYS) SSE
Effluent SS (g/m^3)
Time (days)
494.00000
447.00000
399.00000
349.00000
300.00000
251.00000
202.00000
153.00000
104.00000
55.00000
6.00000
0.00
18.30
35.60
54.90
73.20
91.50
109.80
128.10
146.40
164.70
183.00
201.30
219.60
237.90
256.20
274.50
292.80
311.10
329.40
347.70
366.00

APPENDIX B

EXAMINATION OF THE DISTRIBUTION OF 7-DAY and 30-DAY AVERAGES

To examine the distribution of 7- and 30-day averages of effluent BOD and SS concentrations, a simulation study has been done. From a lognormal population with a mean value of 18 and a variance of 81, 3,150 random variables have been generated. The normal expected frequency of 7-day and 30-day consecutive averages has been examined and shown in Table B-1. The chi-square goodness-of-fit at 5% significant level has been used to test the data for normal distribution. In both cases, the 7-and 30-day averages were not found to be distributed normally.

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA

Table B-1

Simulation study of 7-day and 30-day averages of a lognormal distribution

3150 (Y1, Y2, . . . , Y3150) of lognormal data
with mean 18 and variance 81 has been generated.
N1 is the observed frequency.
NP is the normal expected frequency.

CASE 1: X1 = (Y1 + Y2 + ... + Y7)/7
X2 = (Y8 + Y9 + ... + Y14)/7
.
.
.
X450 = Y3144 + ... + Y3150)/7

INTERVAL	NI	NP
0-12	7	17.50
12-13	22	14.36
13-14	21	22.06
14-15	41	31.09
15-16	41	40.22
16-17	65	47.75
17-18	47	52.02
18-19	50	52.02
19-20	44	47.75
20-21	30	40.22
21-22	18	31.09
22-23	25	22.06
23-24	15	14.36
24-25	5	8.58
25-26	5	4.71
26-27	5	2.37
27-**	9	1.83

The chi-square statistic is: 61.6349
The degree of freedom is: 16

TYPING GUIDE SHEET

CENTER OF PAGE

Table B-1 (continued)

CASE 2:

$$X1 = (Y1 + Y2 + \ldots + Y30)/30$$
$$X2 = (Y31 + Y32 + \ldots + Y60)/30$$
$$\vdots$$
$$X105 = (Y3121 + Y3122 + \ldots + Y3150)/30$$

INTERVAL	NI	NP
0-16	8	11.74
16-17	27	16.76
17-18	28	24.00
18-19	18	24.00
19-20	8	16.76
20-21	7	8.17
21-**	9	3.56

The chi-square statistic is: 22.6486
The degree of freedom is: 6

PAGE NUMBER
TYPING GUIDE SHEET

APPENDIX C

A SHORT DISCUSSION OF THE LOGNORMAL DISTRIBUTION

THE LOGNORMAL DISTRIBUTION - THE MODEL OF PRODUCTS[8]

The normal distribution (or the model of sums) represent those variables which arise as the sum of a number of small random effects. It is desirable to consider the distribution of a phenomenon which arises as the result of the product of a number of variables, e.g., $Y = Y_o W_1 W_2 \ldots . W_n$. Taking the natural logarithms of both sides of the equation above, the result is of the form:

$$\text{Ln } y = \text{Ln } y_o + \text{Ln } W_1 + \text{Ln } W_2 + \ldots \text{Ln } W_n$$

Since W_i are random variables, the functions $\text{Ln}W_i$ are also random variables. Calling upon the central limit theorem which, under very general conditions, says "As the number of variables in the sum becomes large, the distribution of the sum of random variables will approach the normal distribution". This theorem is used to infer that the sum of the logarithm of variables W_i (i.e., $\text{Ln}W_i$) will be approximately normally distributed. Let Ln y = x, knowing that x is normally distributed, our interest is to determine the distribution of y. The random variable y, whose logarithms are normally distributed, is said to have a lognormal distribution, i.e., the random variable y is lognormally distributed, whereas its logarithm x is normally distributed. The relationship between y and x is one-to-one and monotonically related, i.e., the function y=g(u) relating the two random variables is such that it always increases as n increases and is such that there is only a single value of n for each value of y and vice versa. The probability density function of such a case of one-to-one, monotonic transformation is defined as (see reference 8, p. 103-109).

$$fy\ (y) = \left| \frac{dg^{-1}}{dy} (y) \right| \quad fx\ (g^{-1}(y))$$

where

$$y = e^x = g\ (x)$$

$$x = \text{Ln } y = g^{-1}(y)$$

$$\frac{dg^{-1}\ (y)}{dy} = \frac{1}{y}$$

Since x is normally distributed, then

$$f_x\,(n) = \frac{1}{\sigma_x\sqrt{2\pi}}\ \exp\ [\ -\frac{1}{2}\left(\frac{n - m_x}{\sigma_x}\right)^2\ ] \qquad -\infty \le n \le \infty$$

after substituting

$$f_y\,(y) = \frac{1}{y\ \sigma_x\sqrt{2\pi}}\ \exp\ [\ -\frac{1}{2}\left(\frac{\mathrm{Ln}\ y - m_x}{\sigma_x}\right)^2\ ] \qquad y \ge 0$$

In this equation, the range is zero to infinity, whereas the range of x is minus to plus infinity. If y=1, x=0, ad if y > 1, x is positive. In the range of $0 \le y \le 1$, x is in the range of minus infinity to zero, since the logarithm of a number between zero and unity is negative. Y cannot have negative values, since the logarithm of a negative number is not defined.

Knowing that for normal (and other symmetrical) distributions, the median (the median m of a random variable is defined as the value below which one-half of the probability mean lies) equals the mean

$$m_x = m_x$$

or

$$\mathrm{Ln}\ m_y = m_x$$

Substituting for Ln y - m_x,

$$\mathrm{Ln}\ y - m_x = \mathrm{Ln}\ y - \mathrm{Ln}\ m_y = \mathrm{Ln}\ \frac{y}{m_y}$$

also $\sigma_x = \sigma_{\mathrm{Ln}\ y}$

therefore, the probability density function of a lognormal distribution in terms of the more familiar and natural pairs of parameters, i.e., median and standard deviation is

$$f_y\,(y) = \frac{1}{y\sqrt{2\pi}\ \ \sigma_{\mathrm{Lny}}}\ \exp\{\ -\frac{1}{2}\ [\ \frac{1}{\sigma_{\mathrm{Lny}}}\ \mathrm{Ln}\ (\frac{y}{m_y})\ ]\} \qquad y \ge 0$$

APPENDIX D

KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT TEST

The usefulness of a proposed distribution function for modelling BOD and SS data from each plant can be examined using the Kolmogorov-Smirnov (K-S) goodness-of-fit test. The null hypothesis for this test is:

> "The observed data are sampled from an underlying population which is distributed according to the hypothesized distribution function."

It is assumed that if the test hypothesis is true, any differences between the observed distribution and the hypothesized distribution are due to the effects of random sampling.

To perform the K-S test, the observed data are ranked in ascending order and assigned percentile values. The difference, d_i, between the observed percentile and the percentile predicted by the hypothesized distribution is then computed for each data value. The maximum absolute value of d_i is the test statistic, D. A graphical illustration of the computation of D is presented in Figure D-1. The probability, P, that the test hypothesis is true is computed from the value of D and the number of data points, n. (8)

The K-S test is specially useful for comparing a hypothesized distribution with the distribution of a set of data points because the test statistic, D, provides a quantitative measure of the actual percentile difference between the two distributions. The value of D can be thought of as the maximum absolute error that would occur if the hypothesized distribution were used to eliminate the percentiles of the observed data values.

From experience with the K-S test, it can be shown that the largest values of d_i tend to occur between the 30th and 70th percentiles and that the percentile differences between the two distributions are generally smaller in the tails. Therefore, the value of D is typically much larger than the values of d_i for data points in the extremes of the observed distribution.